SECTION 4

QUALITY ASSURANCE

4.1 INTRODUCTION

4.1.1   Development and maintenance of a toxicity test laboratory quality assurance (QA) program (USEPA, 1991b) requires an ongoing commitment by laboratory management. Each toxicity test laboratory should (1) appoint a quality assurance officer with the responsibility and authority to develop and maintain a QA program, (2) prepare a quality assurance plan with stated data quality objectives (DQOs), (3) prepare written descriptions of laboratory standard operating procedures (SOPs) for culturing, toxicity testing, instrument calibration, sample chain-of-custody procedures, laboratory sample tracking system, glassware cleaning, etc., and (4) provide an adequate, qualified technical staff for culturing and toxicity testing the organisms, and suitable space and equipment to assure reliable data.

4.1.2   QA practices for toxicity testing laboratories must address all activities that affect the quality of the final effluent toxicity data, such as: (1) effluent sampling and handling; (2) the source and condition of the test organisms; (3) condition of equipment; (4) test conditions; (5) instrument calibration; (6) replication; (7) use of reference toxicants; (8) record keeping; and (9) data evaluation.

4.1.3   Quality control practices, on the other hand, consist of the more focused, routine, day-to-day activities carried out within the scope of the overall QA program. For more detailed discussion of quality assurance and general guidance on good laboratory practices and laboratory evaluation related to toxicity testing, see FDA (1978); USEPA (1979d); USEPA (1980b); USEPA (1980c); USEPA (1991c); DeWoskin (1984); and Taylor (1987).

4.1.4   Guidelines for the evaluation of laboratory performing toxicity tests and laboratory evaluation criteria are found in USEPA (1991c).

4.2 FACILITIES, EQUIPMENT, AND TEST CHAMBERS

4.2.1   Separate test organism culturing and toxicity testing areas should be provided to avoid possible loss of cultures due to cross-contamination. Ventilation systems should be designed and operated to prevent recirculation or leakage of air from chemical analysis laboratories or sample storage and preparation areas into organism culturing or testing areas, and from testing and sample preparation areas into culture rooms.

4.2.2   Laboratory and toxicity test temperature control equipment must be adequate to maintain recommended test water temperatures. Recommended materials must be used in the fabrication of the test equipment which comes in contact with the effluent (see Section 5, Facilities, Equipment, and Supplies; and specific toxicity test method).

4.3 TEST ORGANISMS

4.3.1   The test organisms used in the procedures described in this manual are the sheepshead minnow, *Cyprinodon variegatus*; the inland silverside, *Menidia beryllina*; the mysid, *Mysidopsis bahia*; the sea urchin, *Arbacia punctulata*; and the red macroalga, *Champia parvula*. The organisms used should be disease-free and appear healthy, behave normally, feed well, and have low mortality in cultures, during holding, and in test control. Test organisms should be positively identified to species (see Section 6, Test Organisms).

4.4 LABORATORY WATER USED FOR CULTURING AND TEST DILUTION WATER

4.4.1   The quality of water used for test organism culturing and for dilution water used in toxicity tests is extremely important. Water for these two uses should come from the same source. The dilution water used in effluent toxicity

tests will depend on the objectives of the study and logistical constraints, as discussed in Section 7, Dilution Water. The dilution water used in the toxicity tests may be natural seawater, hypersaline brine (100‰) prepared from natural seawater, or artificial seawater prepared from commercial sea salts, such as FORTY FATHOMS® or HW MARINEMIX®, if recommended in the method. GP2 synthetic seawater, made from reagent grade chemical salts (30‰) in conjunction with natural seawater, may also be used if recommended. Hypersaline brine and artificial seawater can be used with *Champia parvula* only if they are accompanied by at least 50% natural seawater. Types of water are discussed in Section 5, Facilities, Equipment, and Supplies. Water used for culturing and test dilution water should be analyzed for toxic metals and organics at least annually or whenever difficulty is encountered in meeting minimum acceptability criteria for control survival and reproduction or growth. The concentration of the metals, Al, As, Cr, Co, Cu, Fe, Pb, Ni, Zn, expressed as total metal, should not exceed 1 µg/L each, and Cd, Hg, and Ag, expressed as total metal, should not exceed 100 ng/L each. Total organochlorine pesticides plus PCBs should be less than 50 ng/L (APHA, 1992). Pesticide concentrations should not exceed USEPA's National Ambient Water Quality chronic criteria values where available.

4.5 EFFLUENT AND RECEIVING WATER SAMPLING AND HANDLING

4.5.1 Sample holding times and temperatures of effluent samples collected for on-site and off-site testing must conform to conditions described in Section 8, Effluent and Receiving Water Sampling, Sample Handling, and Sample Preparation for Toxicity Tests.

4.6 TEST CONDITIONS

4.6.1 Water temperature and salinity should be maintained within the limits specified for each test. The temperature of test solutions must be measured by placing the thermometer or probe directly into the test solutions, or by placing the thermometer in equivalent volumes of water in surrogate vessels positioned at appropriate locations among the test vessels. Temperature should be recorded continuously in at least one vessel during the duration of each test. Test solution temperatures should be maintained within the limits specified for each test. DO concentrations and pH should be checked at the beginning of the test and daily throughout the test period.

4.7 QUALITY OF TEST ORGANISMS

4.7.1 The health of test organisms is primarily assessed by the performance (survival, growth, and/or reproduction) of organisms in control treatments of individual tests. The health and sensitivity of test organisms is also assessed by reference toxicant testing. In addition to documenting the sensitivity and health of test organisms, reference toxicant testing is used to initially demonstrate acceptable laboratory performance (Subsection 4.15) and to document ongoing laboratory performance (Subsection 4.16).

4.7.2 Regardless of the source of test organisms (in-house cultures or purchased from external suppliers), the testing laboratory must perform at least one acceptable reference toxicant test per month for each toxicity test method conducted in that month (Subsection 4.16). If a test method is conducted only monthly, or less frequently, a reference toxicant test must be performed concurrently with each effluent toxicity test.

4.7.3 When acute or short-term chronic toxicity tests are performed with effluents or receiving waters using test organisms obtained from outside the test laboratory, concurrent toxicity tests of the same type must be performed with a reference toxicant, unless the test organism supplier provides control chart data from at least the last five monthly short-term chronic toxicity tests using the same reference toxicant and test conditions (see Section 6, Test Organisms).

4.7.4 The supplier should certify the species identification of the test organisms, and provide the taxonomic reference (citation and page) or name(s) of the taxonomic expert(s) consulted.

4.7.5 If a routine reference toxicant test fails to meet test acceptability criteria, then the reference toxicant test must be immediately repeated.

4.8   FOOD QUALITY

4.8.1   The nutritional quality of the food used in culturing and testing fish and invertebrates is an important factor in the quality of the toxicity test data. This is especially true for the unsaturated fatty acid content of brine shrimp nauplii, *Artemia*. Problems with the nutritional suitability of the food will be reflected in the survival, growth, and reproduction of the test organisms in cultures and toxicity tests. *Artemia* cysts and other foods must be obtained as described in Section 5, Facilities, Equipment, and Supplies.

4.8.2   Problems with the nutritional suitability of food will be reflected in the survival, growth, and reproduction of the test organisms in cultures and toxicity tests. If a batch of food is suspected to be defective, the performance of organisms fed with the new food can be compared with the performance of organisms fed with a food of known quality in side-by-side tests. If the food is used for culturing, its suitability should be determined using a short-term chronic test which will determine the affect of food quality on growth or reproduction of each of the relevant test species in culture, using four replicates with each food source. Where applicable, foods used only in chronic toxicity tests can be compared with a food of known quality in side-by-side, multi-concentration chronic tests, using the reference toxicant regularly employed in the laboratory QA program.

4.8.3   New batches of food used in culturing and testing should be analyzed for toxic organics and metals or whenever difficulty is encountered in meeting minimum acceptability criteria for control survival and reproduction or growth. If the concentration of total organochlorine pesticides exceeds 0.15 µg/g wet weight, or the concentration of total organochlorine pesticides plus PCBs exceeds 0.30 µg/g wet weight, or toxic metals (Al, As, Cr, Cd, Cu, Pb, Ni, Zn, expressed as total metal) exceed 20 µg/g wet weight, the food should not be used (for analytical methods, see AOAC, 1990; and USDA, 1989).

4.84   For foods (e.g., YCT) which are used to culture and test organisms, the quality of the food should meet the requirements for the laboratory water used for culturing and test dilution water as described in Section 4.4 above.

4.9   ACCEPTABILITY OF CHRONIC TOXICITY TESTS

4.9.1   The results of the sheepshead minnow, *Cyprinodon variegatus*, inland silverside, *Menidia beryllina*, or mysid, *Mysidopsis bahia*, tests are acceptable if survival in the controls is 80% or greater. The sea urchin, *Arbacia punctulata*, test requires control egg fertilization equal to or exceeding 70%. However, greater than 90% fertilization may result in masking toxic responses. The red macroalga, *Champia parvula*, test is acceptable if survival is 100%, and the mean number of cystocarps per plant should equal or exceed 10. If the sheepshead minnow, *Cyprindon variegatus*, larval survival and growth test is begun with less-than-24-h old larvae, the mean dry weight of the surviving larvae in the control chambers at the end of the test must equal or exceed 0.60 mg, if the weights are determined immediately, or 0.50 mg if the larvae are preserved in a 4% formalin or 70% ethanol solution. If the inland silverside, *Menidia beryllina*, larval survival and growth test is begun with larvae seven days old, the mean dry weight of the surviving larvae in the control chambers at the end of the test must equal or exceed 0.50 mg, if the weights are determined immediately, or 0.43 mg if the larvae are preserved in a 4% formalin or 70% ethanol solution. The mean mysid dry weight of survivors must be at least 0.20 mg. Automatic or hourly feeding will generally provide control mysids with a dry weight of 0.30 mg. At least 50% of the females should bear eggs at the end of the test, but mysid fecundity is not a factor in test acceptability. However, fecundity must equal or exceed 50% to be used as an endpoint in the test. If these criteria are not met, the test must be repeated.

4.9.2   An individual test may be conditionally acceptable if temperature, DO, and other specified conditions fall outside specifications, depending on the degree of the departure and the objectives of the tests (see test conditions and test acceptability criteria summaries). The acceptability of the test will depend on the experience and professional judgment of the laboratory investigator and the reviewing staff of the regulatory authority. Any deviation from test specifications must be noted when reporting data from a test.

4.10 ANALYTICAL METHODS

4.10.1   Routine chemical and physical analyses for culture and dilution water, food, and test solutions must include established quality assurance practices outlined in USEPA methods manuals (USEPA, 1979a and USEPA, 1979b).

4.10.2   Reagent containers should be dated and catalogued when received from the supplier, and the shelf life should not be exceeded. Also, working solutions should be dated when prepared, and the recommended shelf life should be observed.

4.11 CALIBRATION AND STANDARDIZATION

4.11.1   Instruments used for routine measurements of chemical and physical parameters, such as pH, DO, temperature, conductivity, and salinity, must be calibrated and standardized according to instrument manufacturers procedures as indicated in the general section on quality assurance (see USEPA Methods 150.1, 360.1, 170.1, and 120.1 in USEPA, 1979b). Calibration data are recorded in a permanent log book.

4.11.2   Wet chemical methods used to measure hardness, alkalinity, and total residual chlorine, must be standardized prior to use each day according to the procedures for those specific USEPA methods (see USEPA Methods 130.2 and 310.1 in USEPA, 1979b).

4.12 REPLICATION AND TEST SENSITIVITY

4.12.1   The sensitivity of the tests will depend in part on the number of replicates per concentration, the significance level selected, and the type of statistical analysis. If the variability remains constant, the sensitivity of the test will increase as the number of replicates is increased. The minimum recommended number of replicates varies with the objectives of the test and the statistical method used for analysis of the data.

4.13 VARIABILITY IN TOXICITY TEST RESULTS

4.13.1   Factors which can affect test success and precision include: (1) the experience and skill of the laboratory analyst; (2) test organism age, condition, and sensitivity; (3) dilution water quality; (4) temperature control; (5) and the quality and quantity of food provided. The results will depend upon the species used and the strain or source of the test organisms, and test conditions, such as temperature, DO, food, and water quality. The repeatability or precision of toxicity tests is also a function of the number of test organisms used at each toxicant concentration. Jensen (1972) discussed the relationship between sample size (number of fish) and the standard error of the test, and considered 20 fish per concentration as optimum for Probit Analysis.

4.14 TEST PRECISION

4.14.1   The ability of the laboratory personnel to obtain consistent, precise results must be demonstrated with reference toxicants before they attempt to measure effluent toxicity. The single-laboratory precision of each type of test to be used in a laboratory should be determined by performing at least five or more tests with a reference toxicant.

4.14.2   Test precision can be estimated by using the same strain of organisms under the same test conditions, and employing a known toxicant, such as a reference toxicant.

4.14.3   Interlaboratory precision data from a 1991 study of chronic toxicity tests using two reference toxicants with the mysid, *Mysidopsis bahia*, and the inland silverside, *Menidia beryllina*, is listed in Table 1. Table 2 shows interlaboratory precision data from a study of three chronic toxicity test methods using effluent, receiving water, and reference toxicant sample types (USEPA, 2001a; USEPA, 2001b). For the *Mysidopsis bahia* and the *Cyprinodon variegatus* test methods, the effluent sample was a municipal wastewater spiked with KCl, the receiving water sample was a river water spiked with KCl, and the reference toxicant sample was bioassay-grade FORTY FATHOMS®

synthetic seawater spiked with KCl. For the *Menidia beryllina* test method, the effluent sample was an industrial wastewater spiked with $CuSO_4$, the receiving water sample was a natural seawater spiked with $CuSO_4$, and the reference toxicant sample was bioassay-grade FORTY FATHOMS® synthetic seawater spiked with $CuSO_4$. Additional precision data for each of the tests described in this manual are presented in the sections describing the individual test methods.

4.14.4   Additional information on toxicity test precision is provided in the Technical Support Document for Water Quality-based Toxic Control (see pp. 2-4, and 11-15 in USEPA, 1991a).

4.14.5   In cases where the test data are used in Probit Analysis or other point estimation techniques (see Section 9, Chronic Toxicity Test Endpoints and Data Analysis), precision can be described by the mean, standard deviation, and relative standard deviation (percent coefficient of variation, or CV) of the calculated endpoints from the replicated tests. In cases where the test data are used in the Linear Interpolation Method, precision can be estimated by empirical confidence intervals derived by using the ICPIN Method (see Section 9, Chronic Toxicity Test Endpoints and Data Analysis). However, in cases where the results are reported in terms of the No-Observed-Effect-Concentration (NOEC) and Lowest-Observed-Effect-Concentration (LOEC) (see Section 9, Chronic Toxicity Test Endpoints and Data Analysis), precision can only be described by listing the NOEC-LOEC interval for each test. It is not possible to express precision in terms of a commonly used statistic. However, when all tests of the same toxicant yield the same NOEC-LOEC interval, maximum precision has been attained. The "true" no effect concentration could fall anywhere within the interval, NOEC ± (LOEC minus NOEC).

4.14.6   It should be noted here that the dilution factor selected for a test determines the width of the NOEC-LOEC interval and the inherent maximum precision of the test. As the absolute value of the dilution factor decreases, the width of the NOEC-LOEC interval increases, and the inherent maximum precision of the test decreases. When a dilution factor of 0.3 is used, the NOEC could be considered to have a relative uncertainty as high as ± 300%. With a dilution factor of 0.5, the NOEC could be considered to have a relative variability of ± 100%. As a result of the variability of different dilution factors, **USEPA recommends the use of a ≥ 0.5 dilution factor**. Other factors which can affect test precision include: test organism age, condition, and sensitivity; temperature control; and feeding.

TABLE 1.   NATIONAL INTERLABORATORY STUDY OF CHRONIC TOXICITY TEST PRECISION, 1991: SUMMARY OF RESPONSES USING TWO REFERENCE TOXICANTS[1,2]

Organism	Endpoint	No. Labs	KCl(mg/L)[4]	SD	CV(%)[3]
Mysidopsis bahia	Survival, NOEC	34	NA	NA	NA
	Growth, IC25	26	480	3.47	28.9
	Growth, IC50	22	656	3.17	19.3
	Growth, NOEC	32	NA	NA	NA
	Fecundity, NOEC	25	NA	NA	NA

Organism	Endpoint	No. Labs	Cu(mg/L)[4]	SD	CV(%)[3]
Menidia beryllina	Survival, NOEC	19	NA	NA	NA
	Growth, IC25	13	0.144	1.56	43.5
	Growth, IC50	12	0.180	1.87	41.6
	Growth, NOEC	17	NA	NA	NA

[1] From a national study of interlaboratory precision of toxicity test data performed in 1991 by the Environmental Monitoring Systems Laboratory-Cincinnati, U.S. Environmental Protection Agency, Cincinnati, OH 45268. Participants included federal, state, and private laboratories engaged in NPDES permit compliance monitoring.
[2] Static renewal test, using 25 ‰ modified GP2 artificial seawater.
[3] Percent coefficient of variation = (standard deviation X 100)/mean.
[4] Expressed as mean.

TABLE 2.   NATIONAL INTERLABORATORY STUDY OF CHRONIC TOXICITY TEST PRECISION, 2000: PRECISION OF RESPONSES USING EFFLUENT, RECEIVING WATER, AND REFERENCE TOXICANT SAMPLE TYPES[1]

Organism	Endpoint	Number of Tests[2]	CV (%)[3]
Cyprinodon variegatus	Growth, IC25	21	10.5
Menidia beryllina	Growth, IC25	30	43.8
Mysidopsis bahia	Growth, IC25	36	41.3

[1] From EPA's WET Interlaboratory Variability Study (USEPA, 2001a; USEPA, 2001b).
[2] Represents the number of valid tests (i.e., those that met test acceptability criteria) that were used in the analysis of precision. Invalid tests were not used.
[3] CVs based on total interlaboratory variability (including both within-laboratory and between-laboratory components of variability) and averaged across sample types. IC25s or IC50s were pooled for all laboratories to calculate the CV for each sample type. The resulting CVs were then averaged across sample types.

4.15 DEMONSTRATING ACCEPTABLE LABORATORY PERFORMANCE

4.15.1   It is a laboratory's responsibility to demonstrate its ability to obtain consistent, precise results with reference toxicants before it performs toxicity tests with effluents for permit compliance purposes. To meet this requirement, the intralaboratory precision, expressed as percent coefficient of variation (CV%), of each type of test to be used in a laboratory should be determined by performing five or more tests with different batches of test organisms, using the same reference toxicant, at the same concentrations, with the same test conditions (i.e., the same test duration, type of dilution water, age of test organisms, feeding, etc.), and same data analysis methods. A reference toxicant concentration series (0.5 or higher) should be selected that will consistently provide partial mortalities at two or more concentrations.

4.16 DOCUMENTING ONGOING LABORATORY PERFORMANCE

4.16.1   Satisfactory laboratory performance is demonstrated by performing at least one acceptable test per month with a reference toxicant for each toxicity test method conducted in the laboratory during that month. For a given test method, successive tests must be performed with the same reference toxicant, at the same concentrations, in the same dilution water, using the same data analysis methods. Precision may vary with the test species, reference toxicant, and type of test. Each laboratory's reference toxicity data will reflect conditions unique to that facility, including dilution water, culturing, and other variables; however, each laboratory's reference toxicity results should reflect good repeatability.

4.16.2   A control chart should be prepared for each combination of reference toxicant, test species, test conditions, and endpoints. Toxicity endpoints from five or six tests are adequate for establishing the control charts. Successive toxicity endpoints (NOECs, IC25s, LC50s, etc.) should be plotted and examined to determine if the results (X_1) are within prescribed limits (Figure 1). The chart should plot logarithm of concentration on the vertical axis against the date of the test or test number on the horizontal axis. The types of control charts illustrated (see USEPA, 1979a) are used to evaluate the cumulative trend of results from a series of samples, thus reference toxicant test results should not be used as a *de facto* criterion for rejection of individual effluent or receiving water tests. For endpoints that are point estimates (LC50s and IC25s), the cumulative mean ($\overline{X}$) and upper and lower control limits ($\pm 2S$) are re-calculated with each successive test result. Endpoints from hypothesis tests (NOEC, NOAEC) from each test are plotted directly on the control chart. The control limits would consist of one concentration interval above and below the concentration representing the central tendency. After two years of data collection, or a minimum of 20 data points, the control chart should be maintained using only the 20 most recent data points.

4.16.3   Laboratories should compare the calculated CV (i.e., standard deviation / mean) of the IC25 for the 20 most recent data points to the distribution of laboratory CVs reported nationally for reference toxicant testing (Table 3-2 in USEPA, 2000b). If the calculated CV exceeds the 75th percentile of CVs reported nationally, the laboratory should use the 75th and 90th percentiles to calculate warning and control limits, respectively, and the laboratory should investigate options for reducing variability. Note: Because NOECs can only be a fixed number of discrete values, the mean, standard deviation, and CV cannot be interpreted and applied in the same way that these descriptive statistics are interpreted and applied for continuous variables such as the IC25 or LC50.

4.16.4   The outliers, which are values falling outside the upper and lower control limits, and trends of increasing or decreasing sensitivity, are readily identified. In the case of endpoints that are point estimates (LC50s and IC25s), at the $P_{0.05}$ probability level, one in 20 tests would be expected to fall outside of the control limits by chance alone. If more than one out of 20 reference toxicant tests fall outside the control limits, the laboratory should investigate sources of variability, take corrective actions to reduce identified sources of variability, and perform an additional reference toxicant test during the same month. Control limits for the NOECs will also be exceeded occasionally, regardless of how well a laboratory performs. In those instances when the laboratory can document the cause for the outlier (e.g., operator error, culture health or test system failure), the outlier should be excluded from the future calculations of the

control limits.   If two or more consecutive tests do not fall within the control limits, the results must be explained and the reference toxicant test must be immediately repeated. Actions taken to correct the problem must be reported.

4.16.5   If the toxicity value from a given test with a reference toxicant fall <u>well</u> outside the expected range for the test organisms when using the standard dilution water and other test conditions, the laboratory should investigate sources of variability, take corrective actions to reduce identified sources of variability, and perform an additional reference toxicant test during the same month. Performance should improve with experience, and the control limits for endpoints that are point estimates should gradually narrow. However, control limits of ± 2S will be exceeded 5% of the time by chance alone, regardless of how well a laboratory performs. Highly proficient laboratories which develop very narrow control limits may be unfairly penalized if a test result which falls just outside the control limits is rejected *de facto*. For this reason, the width of the control limits should be considered in determining whether or not a reference toxicant test result falls "well" outside the expected range. The width of the control limits may be evaluated by comparing the calculated CV (i.e., standard deviation / mean) of the IC25 for the 20 most recent data points to the distribution of laboratory CVs reported nationally for reference toxicant testing (Table 3-2 in USEPA, 2000b). In determining whether or not a reference toxicant test result falls "well" outside the expected range, the result also may be compared with upper and lower bounds for ± 3S, as any result outside these control limits would be expected to occur by chance only 1 out of 100 tests (Environment Canada, 1990). When a result from a reference toxicant test is outside the 99% confidence intervals, the laboratory must conduct an immediate investigation to assess the possible causes for the outlier.

4.16.6   Reference toxicant test results should not be used as a *de facto* criterion for rejection of individual effluent or receiving water tests. Reference toxicant testing is used for evaluating the health and sensitivity of organisms over time and for documenting initial and ongoing laboratory performance. While reference toxicant test results should not be used as a *de facto* criterion for test rejection, effluent and receiving water test results should be reviewed and interpreted in the light of reference toxicant test results. The reviewer should consider the degree to which the reference toxicant test result fell outside of control chart limits, the width of the limits, the direction of the deviation (toward increased test organism sensitivity or toward decreased test organism sensitivity), the test conditions of both the effluent test and the reference toxicant test, and the objective of the test.

4.17   REFERENCE TOXICANTS

4.17.1   Reference toxicants such as sodium chloride (NaCl), potassium chloride (KCl), cadmium chloride ($CdCl_2$), copper sulfate ($CuSO_4$), sodium dodecyl sulfate (SDS), and potassium dichromate ($K_2Cr_2O_7$), are suitable for use in the NPDES Program and other Agency programs requiring aquatic toxicity tests. EMSL-Cincinnati plans to release USEPA-certified solutions of cadmium and copper for use as reference toxicants, through cooperative research and development agreements with commercial suppliers, and will continue to develop additional reference toxicants for future release.   Standard reference materials can be obtained from commercial supply houses, or can be prepared inhouse using reagent grade chemicals. The regulatory agency should be consulted before reference toxicant(s) are selected and used.
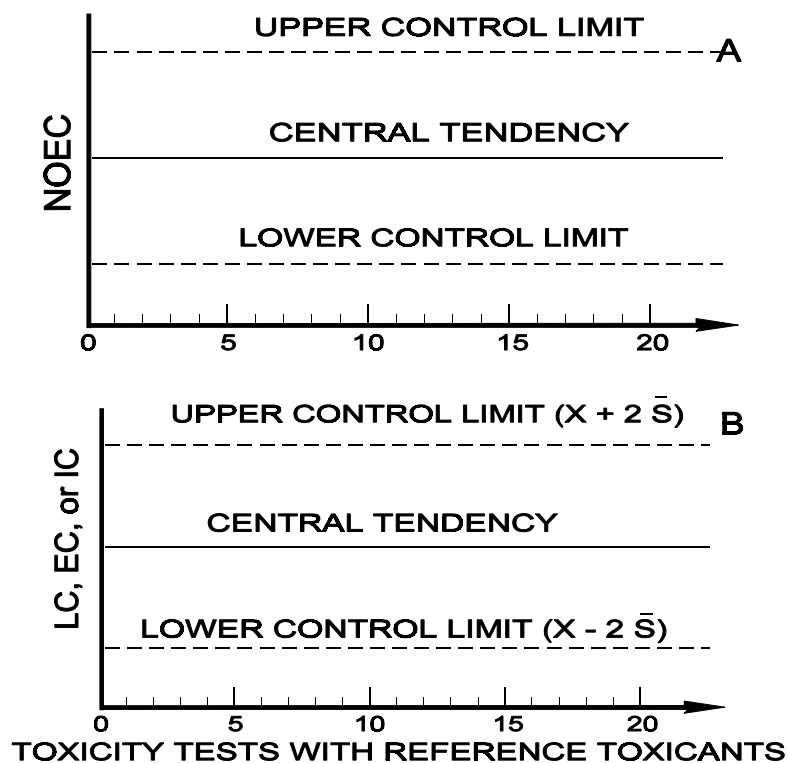
$$\overline{X} = \frac{\sum_{i+1}^{n} X_i}{n}$$

$$S = \frac{\sqrt{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}}{n-1}$$

Where: X_i = Successive toxicity values from toxicity tests.

n = Number of tests.

$\overline{X}$ = Mean toxicity value.

S = Standard deviation.

Figure 1.   Control charts. (A) hypothesis testing results; (B) point estimates (LC, EC, or IC).

4.18 RECORD KEEPING

4.18.1   Proper record keeping is important. A complete file must be maintained for each individual toxicity test or group of tests on closely related samples. This file must contain a record of the sample chain-of-custody; a copy of the sample log sheet; the original bench sheets for the test organism responses during the toxicity test(s); chemical analysis data on the sample(s); detailed records of the test organisms used in the test(s), such as species, source, age, date of receipt, and other pertinent information relating to their history and health; information on the calibration of equipment and instruments; test conditions employed; and results of reference toxicant tests. Laboratory data should be recorded on a real-time basis to prevent the loss of information or inadvertent introduction of errors into the record. Original data sheets should be signed and dated by the laboratory personnel performing the tests.

4.18.2   The regulatory authority should retain records pertaining to discharge permits. Permittees are required to retain records pertaining to permit applications and compliance for a minimum of 3 years [40 CFR 122.41(j)(2)].

SECTION 5

FACILITIES, EQUIPMENT, AND SUPPLIES

5.1 GENERAL REQUIREMENTS

5.1.1 Effluent toxicity tests may be performed in a fixed or mobile laboratory. Facilities must include equipment for rearing and/or holding organisms. Culturing facilities for test organisms may be desirable in fixed laboratories which perform large numbers of tests. Temperature control can be achieved using circulating water baths, heat exchangers, or environmental chambers. Water used for rearing, holding, acclimating, and testing organisms may be natural seawater or water made up from hypersaline brine derived from natural seawater, or water made up from reagent grade chemicals (GP2) or commercial (FORTY FATHOMS® or HW MARINEMIX®) artificial sea salts when specifically recommended in the method. Air used for aeration must be free of oil and toxic vapors. Oil-free air pumps should be used where possible. Particulates can be removed from the air using BALSTON® Grade BX or equivalent filters, and oil and other organic vapors can be removed using activated carbon filters (BALSTON®, C-1 filter, or equivalent).

5.1.2 The facilities must be well ventilated and free of fumes. Laboratory ventilation systems should be checked to ensure that return air from chemistry laboratories and/or sample handling areas is not circulated to test organism culture rooms or toxicity test rooms, or that air from toxicity test rooms does not contaminate culture areas. Sample preparation, culturing, and toxicity testing areas should be separated to avoid cross-contamination of cultures or toxicity test solutions with toxic fumes. Air pressure differentials between such rooms should not result in a net flow of potentially contaminated air to sensitive areas through open or loosely-fitting doors. Organisms should be shielded from external disturbances.

5.1.3 Materials used for exposure chambers, tubing, etc., which come in contact with the effluent and dilution water, should be carefully chosen. Tempered glass and perfluorocarbon plastics (TEFLON®) should be used whenever possible to minimize sorption and leaching of toxic substances. These materials may be reused following decontamination. Containers made of plastics, such as polyethylene, polypropylene, polyvinyl chloride, TYGON®, etc., may be used as test chambers or to ship, store, and transfer effluents and receiving waters, but they should not be reused unless absolutely necessary, because they might carry over adsorbed toxicants from one test to another, if reused. However, these containers may be repeatedly reused for storing uncontaminated waters such as deionized or laboratory-prepared dilution waters and receiving waters. Glass or disposable polystyrene containers can be used as test chambers. The use of large (≥ 20 L) glass carboys is discouraged for safety reasons.

5.1.4 New plastic products of a type not previously used should be tested for toxicity before initial use by exposing the test organisms in the test system where the material is used. Equipment (pumps, valves, etc.) which cannot be discarded after each use because of cost, must be decontaminated according to the cleaning procedures listed below (see Section 5, Facilities, Equipment, and Supplies, Subsection 5.3.2). Fiberglass, in addition to the previously mentioned materials, can be used for holding, acclimating, and dilution water storage tanks, and in the water delivery system, but once contaminated with pollutants the fiberglass should not be reused. All material should be flushed or rinsed thoroughly with the test media before using in the test.

5.1.5 Copper, galvanized material, rubber, brass, and lead must not come in contact with culturing, holding, acclimation, or dilution water, or with effluent samples and test solutions. Some materials, such as several types of neoprene rubber (commonly used for stoppers) may be toxic and should be tested before use.

5.1.6 Silicone adhesive used to construct glass test chambers absorbs some organochlorine and organophosphorus pesticides, which are difficult to remove. Therefore, as little of the adhesive as possible should be in contact with water. Extra beads of adhesive inside the containers should be removed.

5.2 TEST CHAMBERS

5.2.1   Test chamber size and shape are varied according to size of the test organism. Requirements are specified in each toxicity test method.

5.3 CLEANING TEST CHAMBERS AND LABORATORY APPARATUS

5.3.1   New plasticware used for sample collection or organism exposure vessels generally does not require thorough cleaning before use. It is sufficient to rinse new sample containers once with dilution water before use. New, disposable, plastic test chambers may have to be rinsed with dilution water before use.   New glassware must be soaked overnight in 10% acid (see below) and also should be rinsed well in deionized water and seawater.

5.3.2   All non-disposable sample containers, test vessels, pumps, tanks, and other equipment that has come in contact with effluent must be washed after use to remove surface contaminants, as described below.

   1.   Soak 15 minutes in tap water and scrub with detergent, or clean in an automatic dishwasher.
   2.   Rinse twice with tap water.
   3.   Carefully rinse once with fresh dilute (10% V:V) hydrochloric acid or nitric acid to remove
          scale, metals and bases. To prepare a 10% solution of acid, add 10 mL of concentrated acid to
          90 mL of deionized water.
   4.   Rinse twice with deionized water.
   5.   Rinse once with full-strength, pesticide-grade acetone to remove organic compounds (use a
          fume hood or canopy).
   6.   Rinse three times with deionized water.

5.3.3   All test chambers and equipment must be thoroughly rinsed with the dilution water immediately prior to use in each test.

5.4 APPARATUS AND EQUIPMENT FOR CULTURING AND TOXICITY TESTS

5.4.1   Apparatus and equipment requirements for culturing and toxicity tests are specified in each toxicity test method. Also, see USEPA, 2002a.

5.4.2   WATER PURIFICATION SYSTEM

5.4.2.1   A good quality, laboratory grade deionized water, providing a resistance of 18 megaohm-cm, must be available in the laboratory and in sufficient quantity for laboratory needs. Deionized water may be obtained from MILLIPORE®, MILLI-Q®, MILLIPORE® QPAK™$_2$ or equivalent system. If large quantities of high quality deionized water are needed, it may be advisable to supply the laboratory grade water deionizer with preconditioned water from a Culligan®, Continental®, or equivalent mixed-bed water treatment system.

5.5 REAGENTS AND CONSUMABLE MATERIALS

5.5.1   SOURCES OF FOOD FOR CULTURE AND TOXICITY TESTS

   1.   Brine Shrimp, *Artemia* sp. cysts -- Many commercial sources of brine shrimp cysts are available.
   2.   Frozen Adult Brine Shrimp, *Artemia* -- Available from most pet supply shops or other commercial
          sources.
   3.   Flake Food -- TETRAMIN® and BIORIL® or equivalent are available at most pet supply shops.
   4.   Feeding requirements and other specific foods are indicated in the specific toxicity test method.

5.5.1.1   All food should be tested for nutritional suitability and chemically analyzed for organochlorine pesticides, PCBs, and toxic metals (see Section 4, Quality Assurance).

5.5.2 Reagents and consumable materials are specified in each toxicity test method. Also, see Section 4, Quality Assurance.

5.6 TEST ORGANISMS

5.6.1 Test organisms are obtained from inhouse cultures or commercial suppliers (see specific toxicity test method; Sections 4, Quality Assurance and 6, Test Organisms).

5.7 SUPPLIES

5.7.1 See toxicity test methods (see Sections 11-16) for specific supplies.

SECTION 6

TEST ORGANISMS

6.1 TEST SPECIES

6.1.1   The species used in characterizing the chronic toxicity of effluents and/or receiving waters will depend on the requirements of the regulatory authority and the objectives of the test. It is essential that good quality test organisms be readily available throughout the year from inhouse or commercial sources to meet NPDES monitoring requirements. The organisms used in toxicity tests must be identified to species. If there is any doubt as to the identity of the test organisms, representative specimens should be sent to a taxonomic expert to confirm the identification.

6.1.2   Toxicity test conditions and culture methods for the species listed in Subsection 6.1.3 are provided in this manual (also, see USEPA, 2002a).

6.1.3   The organisms used in the short-term tests described in this manual are the sheepshead minnow, *Cyprinodon variegatus*; the inland silverside, *Menidia beryllina*; the mysid, *Mysidopsis bahia*; the sea urchin, *Arbacia punctulata*; and the red macroalga, *Champia parvula*.

6.1.4   Some states have developed culturing and testing methods for indigenous species that may be as sensitive or more sensitive, than the species recommended in Subsection 6.1.3. However, USEPA allows the use of indigenous species only where state regulations require their use or prohibit importation of the species in Subsection 6.1.3. Where state regulations prohibit importation of non-native fishes or use of the recommended test species, permission must be requested from the appropriate state agency prior to their use.

6.1.5   Where states have developed culturing and testing methods for indigenous species other than those recommended in this manual, data comparing the sensitivity of the substitute species and one or more of the recommended species must be obtained in side-by-side toxicity tests with reference toxicants and/or effluents, to ensure that the species selected are at least as sensitive as the recommended species. These data must be submitted to the permitting authority (State or Region) if required. USEPA acknowledges that reference toxicants prepared from pure chemicals may not always be representative of effluents. However, because of the observed and/or potential variability in the quality and toxicity of effluents, it is not possible to specify a representative effluent.

6.1.6   Guidance for the selection of test organisms where the salinity of the effluent and/or receiving water requires special consideration is provided in the Technical Support Document for Water Quality-based Toxics Control (USEPA, 1991a).

1. Where the salinity of the receiving water is < 1‰, freshwater organisms are used regardless of the salinity of the effluent.
2. Where the salinity of the receiving water is ≥1‰, the choice of organisms depends on state water quality standards and/or permit requirements.

6.2 SOURCES OF TEST ORGANISMS

6.2.1   The test organisms recommended in this manual can be cultured in the laboratory using culturing and handling methods for each organism described in the respective test method sections. Also, see USEPA (2002a).

6.2.2   Inhouse cultures should be established wherever it is cost effective. If inhouse cultures cannot be maintained or it is not cost effective, test organisms should be purchased from experienced commercial suppliers (see USEPA, 1993b).

6.2.3   Sheepshead minnows, inland silversides, mysids, and sea urchins may be purchased from commercial suppliers. However, some of these organisms (e.g., adult sheepshead minnows or adult inland silversides) may not always be available from commercial suppliers and may have to be collected in the field and brought back to the laboratory for spawning to obtain eggs and larvae.

6.2.4   If, because of their source, there is any uncertainty concerning the identity of the organisms, it is advisable to have them examined by a taxonomic specialist to confirm their identification. For detailed guidance on identification, see the individual toxicity test methods.

6.2.5   FERAL (NATURAL OCCURRING, WILD CAUGHT) ORGANISMS

6.2.5.1   The use of test organisms taken from the receiving water has strong appeal, and would seem to be the logical approach. However, it is generally impractical and not recommended for the following reasons:

1. Sensitive organisms may not be present in the receiving water because of previous exposure to the effluent or other pollutants.
2. It is often difficult to collect organisms of the required age and quality from the receiving water.
3. Most states require collection permits, which may be difficult to obtain. Therefore, it is usually more cost effective to culture the organisms in the laboratory or obtain them from private, state, or Federal sources. Fish such as sheepshead minnows and silversides, and invertebrates such as mysids, are easily reared in the laboratory or purchased.
4. The required QA/QC records, such as the single-laboratory precision data, would not be available.
5. Since it is mandatory that the identity of test organisms is known to the species level, it would be necessary to examine each organism caught in the wild to confirm its identity, which would usually be impractical or, at the least, very stressful to the organisms.
6. Test organisms obtained from the wild must be observed in the laboratory for a minimum of one week prior to use, to ensure that they are free of signs of parasitic or bacterial infections and other adverse effects. Fish captured by electroshocking must not be used in toxicity testing.

6.2.5.2   Guidelines for collection of natural occurring organisms are provided in USEPA (1973); USEPA (1990a); and USEPA (1993b).

6.2.6   Regardless of their source, test organisms should be carefully observed to ensure that they are free of signs of stress and disease, and in good physical condition. Some species of test organisms, such as trout, can be obtained from stocks certified as "disease-free."

6.3   LIFE STAGE

6.3.1   Young organisms are often more sensitive to toxicants than are adults. For this reason, the use of early life stages, such as juvenile mysids and larval fish, is required for all tests. In a given test, all organisms should be approximately the same age and should be taken from the same source. Since age may affect the results of the tests, it would enhance the value and comparability of the data if the same species in the same life stages were used throughout a monitoring program at a given facility.

6.4   LABORATORY CULTURING

6.4.1   Instructions for culturing and/or holding the recommended test organisms are included in specified test methods (also, see USEPA, 2002a).

6.5 HOLDING AND HANDLING TEST ORGANISMS

6.5.1   Test organisms should not be subjected to changes of more than 3°C in water temperature or 3‰ in salinity in any 12 h period.

6.5.2   Organisms should be handled as little as possible. When handling is necessary, it should be done as gently, carefully, and quickly as possible to minimize stress. Organisms that are dropped or touch dry surfaces or are injured during handling must be discarded. Dipnets are best for handling larger organisms. These nets are commercially available or can be made from small-mesh nylon netting, silk bolting cloth, plankton netting, or similar material. Wide-bore, smooth glass tubes (4 to 8 mm ID) with rubber bulbs or pipettors (such as a PROPIPETTE® or other pipettor) should be used for transferring smaller organisms such as mysids, and larval fish.

6.5.3   Holding tanks for fish are supplied with a good quality water (see Section 5, Facilities, Equipment, and Supplies) with a flow-through rate of at least two tank-volumes per day. Otherwise, use a recirculation system where the water flows through an activated carbon or undergravel filter to remove dissolved metabolites. Culture water can also be piped through high intensity ultraviolet light sources for disinfection, and to photo-degrade dissolved organics.

6.5.4   Crowding should be avoided because it will stress the organisms and lower the DO concentrations to unacceptable levels. The DO must be maintained at a minimum of 4.0 mg/L. The solubility of oxygen depends on temperature, salinity, and altitude. Aerate gently if necessary.

6.5.5   The organisms should be observed carefully each day for signs of disease, stress, physical damage, or mortality. Dead and abnormal organisms should be removed as soon as observed. It is not uncommon for some fish mortality (5-10%) to occur during the first 48 h in a holding tank because of individuals that refuse to feed on artificial food and die of starvation. Organisms in the holding tanks should generally be fed as in the cultures (see culturing methods in the respective methods).

6.5.6   Fish should be fed as much as they will eat at least once a day with live brine shrimp nauplii, *Artemia*, or frozen adult brine shrimp or dry food (frozen food should be completely thawed before use). Adult brine shrimp can be supplemented with commercially prepared food such as TETRAMIN® or BIORIL® flake food, or equivalent. Excess food and fecal material should be removed from the bottom of the tanks at least twice a week by siphoning.

6.5.7   A daily record of feeding, behavioral observations, and mortality should be maintained.

6.6 TRANSPORTATION TO THE TEST SITE

6.6.1   Organisms are transported from the base or supply laboratory to a remote test site in culture water or standard dilution water in plastic bags or large-mouth screw-cap (500 mL) plastic bottles in styrofoam coolers. Adequate DO is maintained by replacing the air above the water in the bags with oxygen from a compressed gas cylinder, and sealing the bags. Another method commonly used to maintain sufficient DO during shipment is to aerate with an airstone which is supplied from a portable pump. The DO concentration must not fall below 4.0 mg/L.

6.6.2   Upon arrival at the test site, organisms are transferred to receiving water if receiving water is to be used as the test dilution water. All but a small volume of the holding water (approximately 5%) is removed by siphoning, and replaced slowly over a 10 to 15 minute period with dilution water. If receiving water is used as dilution water, caution must be exercised in exposing the test organisms to it, because of the possibility that it might be toxic. For this reason, it is recommended that only approximately 10% of the test organisms be exposed initially to the dilution water. If this group does not show excessive mortality or obvious signs of stress in a few hours, the remainder of the test organisms are transferred to the dilution water.

6.6.3   A group of organisms must not be used for a test if they appear to be unhealthy, discolored, or otherwise stressed, or if mortality appears to exceed 10% preceding the test. If the organisms fail to meet these criteria, the entire group must be discarded and a new group obtained. The mortality may be due to the presence of toxicity, if receiving

water is used as dilution water, rather than a diseased condition of the test organisms. If the acclimation process is repeated with a new group of test organisms and excessive mortality occurs, it is recommended that an alternative source of dilution water be used.

6.6.4   The marine organisms can be used at all concentrations of effluent by adjusting the salinity of the effluent to salinities specified for the appropriate species test condition or to the salinity approximating that of the receiving water, by adding sufficient dry ocean salts, such as FORTY FATHOMS®, or equivalent, GP2, or hypersaline brine.

6.6.5   Saline dilution water can be prepared with deionized water or a freshwater such as well water or a suitable surface water. If dry ocean salts are used, care must be taken to ensure that the added salts are completely dissolved and the solution is aerated 24 h before the test organisms are placed in the solutions. The test organisms should be acclimated in synthetic saline water prepared with the dry salts. **Caution**: addition of dry ocean salts to dilution water may result in an increase in pH. (The pH of estuarine and coastal saline waters is normally 7.5-8.3).

6.6.6   All effluent concentrations and the control(s) used in a test should have the same salinity. The change in salinity upon acclimation at the desired test dilution should not exceed 6‰. The required salinities for culturing and toxicity tests with estuarine and marine species are listed in the test method sections.

6.7 TEST ORGANISM DISPOSAL

6.7.1 When the toxicity test(s) is concluded, all test organisms (including controls) should be humanely destroyed and disposed of in an appropriate manner.

SECTION 7

DILUTION WATER

7.1 TYPES OF DILUTION WATER

7.1.1 The type of dilution water used in effluent toxicity tests will depend largely on the objectives of the study.

7.1.1.1 If the objective of the test is to estimate the absolute chronic toxicity of the effluent, a synthetic (standard) dilution water is used. If the test organisms have been cultured in water which is different from the test dilution water, a second set of controls, using culture water, should be included in the test.

7.1.1.2 If the objective of the test is to estimate the chronic toxicity of the effluent in uncontaminated receiving water, the test may be conducted using dilution water consisting of a single grab sample of receiving water (if non-toxic), collected outside the influence of the outfall, or with other uncontaminated natural water (surface water) or standard dilution water having approximately the same salinity as the receiving water. Seasonal variations in the quality of receiving waters may affect effluent toxicity. Therefore, the salinity of saline receiving water samples should be determined before each use. If the test organisms have been cultured in water which is different from the test dilution water, a second set of controls, using culture water, should be included in the test.

7.1.1.3 If the objective of the test is to determine the additive or mitigating effects of the discharge on already contaminated receiving water, the test is performed using dilution water consisting of receiving water collected outside the influence of the outfall. A second set of controls, using culture water, should be included in the test.

7.1.2 An acceptable dilution water is one which is appropriate for the objectives of the test; supports adequate performance of the test organisms with respect to survival, growth, reproduction, or other responses that may be measured in the test (i.e., consistently meets test acceptability criteria for control responses); is consistent in quality; and does not contain contaminants that could produce toxicity. Receiving waters, synthetic waters, or synthetic waters adjusted to approximate receiving water characteristics may be used for dilution provided that the water meets the above listed qualifications for an acceptable dilution water. USEPA (2000a) provides additional guidance on selecting appropriate dilution waters.

7.1.3 When dual controls (one control using culture water and one control using dilution water) are used (see Subsections 7.1.1.1 - 7.1.1.3 above), the dilution water control should be used to determine test acceptability. It is also the dilution water control that should be compared to effluent treatments in the calculation and reporting of test results. The culture water control should be used to evaluate the appropriateness of the dilution water source. Significant differences between organism responses in culture water and dilution water controls could indicate toxicity in the dilution water and may suggest an alternative dilution water source. USEPA (2000a) provides additional guidance on dual controls.

7.2 STANDARD, SYNTHETIC DILUTION WATER

7.2.1 Standard, synthetic, dilution water is prepared with deionized water and reagent grade chemicals (GP2) or commercial sea salts (FORTY FATHOMS®, HW MARINEMIX®) (Table 3). The source water for the deionizer can be ground water or tap water.

7.2.2 DEIONIZED WATER USED TO PREPARE STANDARD, SYNTHETIC, DILUTION WATER

7.2.2.1 Deionized water is obtained from a MILLIPORE MILLI-Q®, MILLIPORE® QPAK™$_2$ or equivalent system. It is advisable to provide a preconditioned (deionized) feed water by using a Culligan®, Continental®, or equivalent system in front of the MILLI-Q® System to extend the life of the MILLI-Q® cartridges (see Section 5, Facilities, Equipment, and Supplies).

7.2.2.2   The recommended order of the cartridges in a four-cartridge deionizer (i.e., MILLI-Q® System or equivalent) is: (1) ion exchange, (2) ion exchange, (3) carbon, and (4) organic cleanup (such as ORGANEX-Q®, or equivalent), followed by a final bacteria filter. The QPAK™₂ water system is a sealed system which does not allow for the rearranging of the cartridges. However, the final cartridge is an ORGANEX-Q® filter, followed by a final bacteria filter. Commercial laboratories using this system have not experienced any difficulty in using the water for culturing or testing. Reference to the MILLI-Q® systems throughout the remainder of the manual includes all MILLIPORE® or equivalent systems.

7.2.3   STANDARD, SYNTHETIC SEAWATER

7.2.3.1   To prepare 20 L of a standard, synthetic, reconstituted seawater (modified GP2), using reagent grade chemicals (Table 3), with a salinity of 31‰, follow the instructions below. Other salinities can be prepared by making the appropriate dilutions. Larger or smaller volumes of modified GP2 can be prepared by using proportionately larger or smaller amounts of salts and dilution water.

1.   Place 20 L of MILLI-Q® or equivalent deionized water in a properly cleaned plastic carboy.
2.   Weigh reagent grade salts listed in Table 3 and add, one at a time, to the deionized water. Stir well after adding each salt.
3.   Aerate the final solution at a rate of 1 L/h for 24 h.
4.   Check the pH and salinity.

7.2.3.2   Synthetic seawater can also be prepared by adding commercial sea salts, such as FORTY FATHOMS®, HW MARINEMIX®, or equivalent, to deionized water. For example, thirty-one parts per thousand (31‰) FORTY FATHOMS® can be prepared by dissolving 31 g of sea salts per liter of deionized water. The salinity of the resulting solutions should be checked with a refractometer.

7.2.4   Artificial seawater is to be used only if specified in the method. EMSL-Cincinnati has found FORTY FATHOMS® artificial sea salts suitable for maintaining and spawning the sheepshead minnow, *Cyprinodon variegatus*, and for its use in the sheepshead minnow larval survival and growth test, suitable for maintaining and spawning the inland silverside, *Menidia beryllina*, and for its use in the inland silverside larval survival and growth test, suitable for culturing and maintaining mysid shrimp, *Mysidopsis bahia*, and its use in the mysid shrimp survival, growth, and fecundity test, and suitable for maintaining sea urchins, *Arbacia punctulata*, and for its use in the sea urchin fertilization test. The USEPA Region 6 Houston Laboratory has successfully used HW MARINEMIX® sea salts to maintain and spawn sheepshead minnows, and perform the larval survival and growth test and the embryo-larval survival and teratogenicity test. Also, HW MARINEMIX® sea salts has been used successfully to culture and maintain the mysid brood stock and perform the mysid survival, growth, fecundity test. An artificial seawater formulation, GP2 (Spotte et al., 1984), Table 3, has been used by the Environmental Research Laboratory-Narragansett, RI for all but the embryo-larval survival and teratogenicity test. The suitability of GP2 as a medium for culturing organisms has not been determined.

TABLE 3. PREPARATION OF GP2 ARTIFICIAL SEAWATER USING REAGENT GRADE CHEMICALS[1,2,3]

Compound	Concentration (g/L)	Amount (g) Required for 20 L
NaCl	21.03	420.6
Na_2SO_4	3.52	70.4
KCl	0.61	12.2
KBr	0.088	1.76
$Na_2B_4O_7 \cdot 10\ H_2O$	0.034	0.68
$MgCl_2 \cdot 6\ H_2O$	9.50	190.0
$CaCl_2 \cdot 2\ H_2O$	1.32	26.4
$SrCl_2 \cdot 6\ H_2O$	0.02	0.400
$NaHCO_3$	0.17	3.40

[1] Modified GP2 from Spotte et al. (1984).
[2] The constituent salts and concentrations were taken from USEPA (2002a). The salinity is 30.89 g/L.
[3] GP2 can be diluted with deionized (DI) water to the desired test salinity.

7.3 USE OF RECEIVING WATER AS DILUTION WATER

7.3.1 If the objectives of the test require the use of uncontaminated receiving water as dilution water, and the receiving water is uncontaminated, it may be possible to collect a sample of the receiving water close to the outfall, but should be away from or beyond the influence of the effluent. However, if the receiving water is contaminated, it may be necessary to collect the sample in an area "remote" from the discharge site, matching as closely as possible the physical and chemical characteristics of the receiving water near the outfall.

7.3.2 The sample should be collected immediately prior to the test, but never more than 96 h before the test begins. Except where it is used within 24 h, or in the case where large volumes are required for flow through tests, the sample should be chilled to 0-6°C during or immediately following collection, and maintained at that temperature prior to use in the test.

7.3.3 The investigator should collect uncontaminated water having a salinity as near as possible to the salinity of the receiving water at the discharge site. Water should be collected at slack high tide, or within one hour after high tide. If there is reason to suspect contamination of the water in the estuary, it is advisable to collect uncontaminated water from an adjacent estuary. At times it may be necessary to collect water at a location closer to the open sea, where the salinity is relatively high. In such cases, deionized water or uncontaminated freshwater is added to the saline water to dilute it to the required test salinity. Where necessary, the salinity of a surface water can be increased by the addition of artificial sea salts, such as FORTY FATHOMS®, HW MARINEMIX®, or equivalent, GP2, a

natural seawater of higher salinity, or hypersaline brine. Instructions for the preparation of hypersaline brine by concentrating natural seawater are provided below.

7.3.4 Receiving water containing debris or indigenous organisms, that may be confused with or attack the test organisms, should be filtered through a sieve having 60 μm mesh openings prior to use.

7.3.5 HYPERSALINE BRINE

7.3.5.1 Hypersaline brine (HSB) has several advantages that make it desirable for use in toxicity testing. It can be made from any high quality, filtered seawater by evaporation, and can be added to deionized water to prepare dilution water, or to effluents or surface waters to increase their salinity.

7.3.5.2 The ideal container for making HSB from natural seawater is one that (l) has a high surface to volume ratio, (2) is made of a noncorrosive material, and (3) is easily cleaned (fiberglass containers are ideal). Special care should be used to prevent any toxic materials from coming in contact with the seawater being used to generate the brine. If a heater is immersed directly into the seawater, ensure that the heater materials do not corrode or leach any substances that would contaminate the brine. One successful method used is a thermostatically controlled heat exchanger made from fiberglass. If aeration is used, use only oil-free air compressors to prevent contamination.

7.3.5.3 Before adding seawater to the brine generator, thoroughly clean the generator, aeration supply tube, heater, and any other materials that will be in direct contact with the brine. A good quality biodegradable detergent should be used, followed by several thorough deionized water rinses. High quality (and preferably high salinity) seawater should be filtered to at least 10 mm before placing into the brine generator. Water should be collected on an incoming tide to minimize the possibility of contamination.

7.3.5.4 The temperature of the seawater is increased slowly to 40°C. The water should be aerated to prevent temperature stratification and to increase water evaporation. The brine should be checked daily (depending on the volume being generated) to ensure that the salinity does not exceed 100‰ and that the temperature does not exceed 40°C. Additional seawater may be added to the brine to obtain the volume of brine required.

7.3.5.5 After the required salinity is attained, the HSB should be filtered a second time through a l-μm filter and poured directly into portable containers (20-L CUBITAINERS® or polycarbonate water cooler jugs are suitable). The containers should be capped and labelled with the date the brine was generated and its salinity. Containers of HSB should be stored in the dark and maintained under room temperature until used.

7.3.5.6 If a source of HSB is available, test solutions can be made by following the directions below. Thoroughly mix together the deionized water and brine before mixing in the effluent.

7.3.5.7 Divide the salinity of the HSB by the expected test salinity to determine the proportion of deionized water to brine. For example, if the salinity of the brine is 100‰ and the test is to be conducted at 25‰, 100‰ divided by 25‰ = 4.0. The proportion of brine is 1 part in 4 (one part brine to three parts deionized water).

7.3.5.8 To make 1 L of seawater at 25‰ salinity from a hypersaline brine of 100‰, 250 mL of brine and 750 mL of deionized water are required.

7.4 USE OF TAP WATER AS DILUTION WATER

7.4.1 The use of tap water in the reconstituting of synthetic (artificial) seawater as dilution water is discouraged unless it is dechlorinated and fully treated. Tap water can be dechlorinated by deionization, carbon filtration, or the use of sodium thiosulfate. Use of 3.6 mg/L (anhydrous) sodium thiosulfate will reduce 1.0 mg chlorine/L (APHA, 1992). Following dechlorination, total residual chlorine should not exceed 0.01 mg/L. Because of the possible

toxicity of thiosulfate to test organisms, a control lacking thiosulfate should be included in toxicity tests utilizing thiosulfate-dechlorinated water.

7.4.2   To be adequate for general laboratory use following dechlorination, the tap water is passed through a deionizer and carbon filter to remove toxic metals and organics, and to control hardness and alkalinity.

7.5   DILUTION WATER HOLDING

7.5.1   A given batch of dilution water should not be used for more than 14 days following preparation because of the possible build up of bacterial, fungal, or algal slime growth and the problems associated with it. The container should be kept covered and the contents should be protected from light.

SECTION 8

EFFLUENT AND RECEIVING WATER SAMPLING, SAMPLE HANDLING, AND SAMPLE PREPARATION FOR TOXICITY TESTS

8.1 EFFLUENT SAMPLING

8.1.1   The effluent sampling point should be the same as that specified in the NPDES discharge permit (USEPA, l988b). Conditions for exception would be: (l) better access to a sampling point between the final treatment and the discharge outfall; (2) if the processed waste is chlorinated prior to discharge, it may also be desirable to take samples prior to contact with the chlorine to determine toxicity of the unchlorinated effluent; or (3) in the event there is a desire to evaluate the toxicity of the influent to municipal waste treatment plants or separate wastewater streams in industrial facilities prior to their being combined with other wastewater streams or non-contact cooling water, additional sampling points may be chosen.

8.1.2   The decision on whether to collect grab or composite samples is based on the objectives of the test and an understanding of the short and long-term operations and schedules of the discharger. If the effluent quality varies considerably with time, which can occur where holding times are short, grab samples may seem preferable because of the ease of collection and the potential of observing peaks (spikes) in toxicity. However, the sampling duration of a grab sample is so short that full characterization of an effluent over a 24-h period would require a prohibitively large number of separate samples and tests. Collection of a 24-h composite sample, however, may dilute toxicity spikes, and average the quality of the effluent over the sampling period. Sampling recommendations are provided below (also see USEPA, 2002a).

8.1.3   Aeration during collection and transfer of effluents should be minimized to reduce the loss of volatile chemicals.

8.1.4   Details of date, time, location, duration, and procedures used for effluent sample and dilution water collection should be recorded.

8.2 EFFLUENT SAMPLE TYPES

8.2.1   The advantages and disadvantages of effluent grab and composite samples are listed below:

8.2.1.1 GRAB SAMPLES

Advantages:

1.   Easy to collect; require a minimum of equipment and on-site time.
2.   Provide a measure of instantaneous toxicity. Toxicity spikes are not masked by dilution.

Disadvantages:

1.   Samples are collected over a very short period of time and on a relatively infrequent basis. The chances of detecting a spike in toxicity would depend on the frequency of sampling, and the probability of missing spikes is high.

8.2.1.2 COMPOSITE SAMPLES:

Advantages:

1.   A single effluent sample is collected over a 24-h period.
2.   The sample is collected over a much longer period of time than grab samples and contains all toxicity spikes.

Disadvantages:

1.   Sampling equipment is more sophisticated and expensive, and must be placed on-site for at least 24 h.
2.   Toxicity spikes may not be detected because they are masked by dilution with less toxic wastes.

8.3   EFFLUENT SAMPLING RECOMMENDATIONS

8.3.1   When tests are conducted on-site, test solutions can be renewed daily with freshly collected samples.

8.3.2   When tests are conducted off-site, a minimum of three samples are collected. If these samples are collected on Test Days 1, 3, and 5, the first sample would be used for test initiation, and for test solution renewal on Day 2. The second sample would be used for test solution renewal on Days 3 and 4. The third sample would be used for test solution renewal on Days 5, 6, and 7.

8.3.3   Sufficient sample must be collected to perform the required toxicity and chemical tests. A 4-L (1-gal) CUBITAINER® will provide sufficient sample volume for most tests.

8.3.4   THE FOLLOWING EFFLUENT SAMPLING METHODS ARE RECOMMENDED:

8.3.4.1   Continuous Discharges

8.3.4.1.1   If the facility discharge is continuous, a single 24-h composite sample is to be taken.

8.3.4.2   Intermittent Discharges

8.3.4.2.1   If the facility discharge is intermittent, a composite sample is to be collected for the duration of the discharge but not more than 24 hours.

8.4   RECEIVING WATER SAMPLING

8.4.1   Logistical problems and difficulty in securing sampling equipment generally preclude the collection of composite receiving water samples for toxicity tests. Therefore, based on the requirements of the test, a single grab sample or daily grab samples of receiving water is collected for use in the test.

8.4.2   The sampling point is determined by the objectives of the test. At estuarine and marine sites, samples should be collected at mid-depth.

8.4.3   To determine the extent of the zone of toxicity in the receiving water at estuarine and marine effluent sites, receiving water samples are collected at several distances away from the discharge. The time required for the effluent-receiving-water mixture to travel to sampling points away from the effluent, and the rate and degree of mixing, may be difficult to ascertain. Therefore, it may not be possible to correlate receiving water toxicity with effluent toxicity at the discharge point unless a dye study is performed. The toxicity of receiving water samples from five stations in the discharge plume can be evaluated using the same number of test vessels and test organisms as used in one effluent toxicity test with five effluent dilutions.

8.5 EFFLUENT AND RECEIVING WATER SAMPLE HANDLING, PRESERVATION, AND SHIPPING

8.5.1   Unless the samples are used in an on-site toxicity test the day of collection (or hand delivered to the testing laboratory for use on the day of collection), it is recommended that they be held at 0-6°C until used to inhibit microbial degradation, chemical transformations, and loss of highly volatile toxic substances.

8.5.2   Composite samples should be chilled as they are collected. Grab samples should be chilled immediately following collection.

8.5.3   If the effluent has been chlorinated, total residual chlorine must be measured immediately following sample collection.

8.5.4   Sample holding time begins when the last grab sample in a series is taken (i.e., when a series of four grab samples are taken over a 24-h period), or when a 24-h composite sampling period is completed. If the data from the samples are to be acceptable for use in the NPDES Program, the lapsed time (holding time) from sample collection to first use of each grab or composite sample must not exceed 36 h. EPA believes that 36 h is adequate time to deliver the sample to the laboratories performing the test in most cases. In the isolated cases, where the permittee can document that this delivery time cannot be met, the permitting authority can allow an option for on-site testing or a variance for an extension of shipped sample holding time. The request for a variance in sample holding time, directed to the USEPA Regional Administrator under 40 CFR 136.3(e), should include supportive data which show that the toxicity of the effluent sample is not reduced (e.g., because of volatilization and/or sorption of toxics on the sample container surfaces) by extending the holding time beyond more than 36 h. However, in no case should more than 72 h elapse between collection and first use of the sample. In static-renewal tests, each grab or composite sample may also be used to prepare test solutions for renewal at 24 h and/or 48 h after first use, if stored at 0-6°C, with minimum head space, as described in Subsection 8.5. If shipping problems (e.g., unsuccessful Saturday delivery) are encountered with renewal samples after a test has been initiated, the permitting authority may allow the continued use of the most recently used sample for test renewal. Guidance for determining the persistence of the sample is provided in Subsection 8.7.

8.5.5   To minimize the loss of toxicity due to volatilization of toxic constituents, all sample containers should be "completely" filled, leaving no air space between the contents and the lid.

8.5.6   SAMPLES USED IN ON-SITE TESTS

8.5.6.1   Samples collected for on-site tests should be used within 24 h.

8.5.7   SAMPLES SHIPPED TO OFF SITE FACILITIES

8.5.7.1   Samples collected for off site toxicity testing are to be chilled to 0-6°C during or immediately after collection, and shipped iced to the performing laboratory. Sufficient ice should be placed with the sample in the shipping container to ensure that ice will still be present when the sample arrives at the laboratory and is unpacked. Insulating material should not be placed between the ice and the sample in the shipping container unless required to prevent breakage of glass sample containers.

8.5.7.2   Samples may be shipped in one or more 4-L (l-gal) CUBITAINERS® or new plastic "milk" jugs. All sample containers should be rinsed with source water before being filled with sample. After use with receiving water or effluents, CUBITAINERS® and plastic jugs are punctured to prevent reuse.

8.5.7.3   Several sample shipping options are available, including Express Mail, air express, bus, and courier service. Express Mail is delivered seven days a week. Saturday and Sunday shipping and receiving schedules of private carriers vary with the carrier.

8.6 SAMPLE RECEIVING

8.6.1   Upon arrival at the laboratory, samples are logged in and the temperature is measured and recorded. If the samples are not immediately prepared for testing, they are stored at 0-6°C until used.

8.6.2   Every effort must be made to initiate the test with an effluent sample on the day of arrival in the laboratory, and the sample holding time should not exceed 36 h unless a variance has been granted by the NPDES permitting authority.

8.7 PERSISTENCE OF EFFLUENT TOXICITY DURING SAMPLE SHIPMENT AND HOLDING

8.7.1   The persistence of the toxicity of an effluent prior to its use in a toxicity test is of interest in assessing the validity of toxicity test data, and in determining the possible effects of allowing an extension of the holding time. Where a variance in holding time (> 36 h, but ≤ 72 h) is requested by a permittee (See Subsection 8.5.4), information on the effects of the extension in holding time on the toxicity of the samples must be obtained by comparing the results of multi-concentration chronic toxicity tests performed on effluent samples held 36 h with toxicity test results using the same samples after they were held for the requested, longer period. The portion of the sample set aside for the second test must be held under the same conditions as during shipment and holding.

8.8 PREPARATION OF EFFLUENT AND RECEIVING WATER SAMPLES FOR TOXICITY TESTS

8.8.1   Adjust the sample salinity to the level appropriate for objectives of the study using hypersaline brine or artificial sea salts.

8.8.2   When aliquots are removed from the sample container, the head space above the remaining sample should be held to a minimum. Air which enters a container upon removal of sample should be expelled by compressing the container before reclosing, if possible (i.e., where a CUBITAINER® used), or by using an appropriate discharge valve (spigot).

8.8.3   It may be necessary to first coarse-filter samples through a NYLON® sieve having 2 to 4 mm mesh openings to remove debris and/or break up large floating or suspended solids. If samples contain indigenous organisms that may attack or be confused with the test organisms, the samples should be filtered through a sieve with 60-µm mesh openings. Since filtering may increase the dissolved oxygen (DO) in an effluent, the DO should be checked both before and after filtering. Low dissolved oxygen concentrations will indicate a potential problem in performing the test. **Caution**: filtration may remove some toxicity.

8.8.4   If the samples must be warmed to bring them to the prescribed test temperature, supersaturation of the dissolved oxygen and nitrogen may become a problem. To avoid this problem, samples may be warmed slowly in open test containers. If DO is still above 100% saturation, based on temperature and salinity (Table 4), after warming to test temperature, samples should be aerated moderately (approximately 500 mL/min) for a few minutes using an airstone. If DO is below 4.0 mg/L, the solutions must be aerated moderately (approximately 500 mL/min) for a few minutes, using an airstone, until the DO is within the prescribed range (≥ 4.0 mg/L). **Caution**: avoid excessive aeration.

8.8.4.1   Aeration during the test may alter the results and should be used only as a last resort to maintain the required DO. Aeration can reduce the apparent toxicity of the test solutions by stripping them of highly volatile toxic substances, or increase their toxicity by altering the pH. However, the DO in the test solution should not be permitted to fall below 4.0 mg/L.

8.8.4.2   In static tests (non-renewal or renewal) low DOs may commonly occur in the higher concentrations of wastewater. Aeration is accomplished by bubbling air through a pipet at the rate of 100 bubbles/min. If aeration is necessary, all test solutions must be aerated. It is advisable to monitor the DO closely during the first few hours of the test. Samples with a potential DO problem generally show a downward trend in DO within 4 to 8 h after the test

is started. Unless aeration is initiated during the first 8 h of the test, the DO may be exhausted during an unattended period, thereby invalidating the test.

8.8.5   At a minimum, pH, conductivity or salinity, and total residual chlorine are measured in the undiluted effluent or receiving water, and pH and conductivity are measured in the dilution water.

8.8.5.1   It is recommended that total alkalinity and total hardness also be measured in the undiluted effluent test water and the dilution water.

8.8.6   Total ammonia is measured in effluent and receiving water samples where toxicity may be contributed by unionized ammonia (i.e., where total ammonia $\geq$ 5 mg/L). The concentration (mg/L) of unionized (free) ammonia in a sample is a function of temperature and pH, and is calculated using the percentage value obtained from Table 5, under the appropriate pH and temperature, and multiplying it by the concentration (mg/L) of total ammonia in the sample.

8.8.7   Effluents and receiving waters can be dechlorinated using 6.7 mg/L anhydrous sodium thiosulfate to reduce 1 mg/L chlorine (APHA, 1992). Note that the amount of thiosulfate required to dechlorinate effluents is greater than the amount needed to dechlorinate tap water, (see Section 7, Dilution Water). Since thiosulfate may contribute to sample toxicity, a thiosulfate control should be used in the test in addition to the normal dilution water control.

8.8.8   The DO concentration in the samples should be near saturation prior to use. Aeration will bring the DO and other gases into equilibrium with air, minimize oxygen demand, and stabilize the pH. However, aeration during collection, transfer, and preparation of samples should be minimized to reduce the loss of volatile chemicals.

8.8.9   Mortality or impairment of growth or reproduction due to pH alone may occur if the pH of the sample falls outside the range of 6.0 - 9.0. Thus, the presence of other forms of toxicity (metals and organics) in the sample may be masked by the toxic effects of low or high pH. The question about the presence of other toxicants can be answered only by performing two parallel tests, one with an adjusted pH, and one without an adjusted pH. Freshwater samples are adjusted to pH 7.0, and marine samples are adjusted to pH 8.0, by adding 1N NaOH or 1N HCl dropwise, as required, being careful to avoid overadjustment.

TABLE 4. OXYGEN SOLUBILITY (MG/L) IN WATER AT EQUILIBRIUM WITH AIR AT 760 MM HG
(AFTER Richards and Corwin, 1956)

TEMP	SALINITY (‰)									
(C°)	0	5	10	15	20	25	30	35	40	43
0	14.2	13.8	13.4	12.9	12.5	12.1	11.7	11.2	10.8	10.6
1	13.8	13.4	13.0	12.6	12.2	11.8	11.4	11.0	10.6	10.3
2	13.4	13.0	12.6	12.2	11.9	11.5	11.1	10.7	10.3	10.0
3	13.1	12.7	12.3	11.9	11.6	11.2	10.8	10.4	10.0	9.8
4	12.7	12.3	12.0	11.6	11.3	10.9	10.5	10.1	9.8	9.5
5	12.4	12.0	11.7	11.3	11.0	10.6	0.2	9.8	9.5	9.3
6	12.1	11.7	11.4	11.0	10.7	10.3	0.0	9.6	9.3	9.1
8	11.5	11.2	10.8	10.5	10.2	9.8	9.5	9.2	8.9	8.7
10	10.9	10.7	10.3	10.0	9.7	9.4	9.1	8.8	8.5	8.3
12	10.5	10.2	9.9	9.6	9.3	9.0	8.7	8.4	8.1	7.9
14	10.0	9.7	9.5	9.2	8.9	8.6	8.3	8.1	7.8	7.6
16	9.6	9.3	9.1	8.8	8.5	8.3	8.0	7.7	7.5	7.3
18	9.2	9.0	8.7	8.5	8.2	8.0	7.7	7.5	7.2	7.1
20	8.9	8.6	8.4	8.1	7.9	7.7	7.4	7.2	6.9	6.8
22	8.6	8.4	8.1	7.9	7.6	7.4	7.2	6.9	6.7	6.6
24	8.3	8.1	7.8	7.6	7.4	7.2	6.9	6.7	6.5	6.4
26	8.1	7.8	7.6	7.4	7.2	7.0	6.7	6.5	6.3	6.1
28	7.8	7.6	7.4	7.2	7.0	6.8	6.5	6.3	6.1	6.0
30	7.6	7.4	7.1	6.9	6.7	6.5	6.3	6.1	5.9	5.8
32	7.3	7.1	6.9	6.7	6.5	6.3	6.1	5.9	5.7	5.6

TABLE 5.   PERCENT UNIONIZED NH_3 IN AQUEOUS AMMONIA SOLUTIONS: TEMPERATURE 15-26°C AND pH 6.0-8.9[1]

pH	TEMPERATURE (°C)											
	15	16	17	18	19	20	21	22	23	24	25	26
6.0	0.0274	0.0295	0.0318	0.0343	0.0369	0.0397	0.0427	0.0459	0.0493	0.0530	0.0568	0.0610
6.1	0.0345	0.0372	0.0400	0.0431	0.0464	0.0500	0.0537	0.0578	0.0621	0.0667	0.0716	0.0768
6.2	0.0434	0.0468	0.0504	0.0543	0.0584	0.0629	0.0676	0.0727	0.0781	0.0901	0.0901	0.0966
6.3	0.0546	0.0589	0.0634	0.0683	0.0736	0.0792	0.0851	0.0915	0.0983	0.1134	0.1134	0.1216
6.4	0.0687	0.0741	0.0799	0.0860	0.0926	0.0996	0.107	0.115	0.124	0.133	0.143	0.153
6.5	0.0865	0.0933	0.1005	0.1083	0.1166	0.1254	0.135	0.145	0.156	0.167	0.180	0.19
6.6	0.109	0.117	0.127	0.136	0.147	0.158	0.170	0.182	0.196	0.210	0.226	0.242
6.7	0.137	0.148	0.159	0.171	0.185	0.199	0.214	0.230	0.247	0.265	0.284	0.305
6.8	0.172	0.186	0.200	0.216	0.232	0.250	0.269	0.289	0.310	0.333	0.358	0.384
6.9	0.217	0.234	0.252	0.271	0.292	0.314	0.338	0.363	0.390	0.419	0.450	0.482
7.0	0.273	0.294	0.317	0.342	0.368	0.396	0.425	0.457	0.491	0.527	0.566	0.607
7.1	0.343	0.370	0.399	0.430	0.462	0.497	0.535	0.575	0.617	0.663	0.711	0.762
7.2	0.432	0.466	0.502	0.540	0.581	0.625	0.672	0.722	0.776	0.833	0.893	0.958
7.3	0.543	0.586	0.631	0.679	0.731	0.786	0.845	0.908	0.975	1.05	1.12	1.20
7.4	0.683	0.736	0.793	0.854	0.918	0.988	1.061	1.140	1.224	1.31	1.41	1.51
7.5	0.858	0.925	0.996	1.07	1.15	1.24	1.33	1.43	1.54	1.65	1.77	1.89
7.6	1.08	1.16	1.25	1.35	1.45	1.56	1.67	1.80	1.93	2.07	2.21	2.37
7.7	1.35	1.46	1.57	1.69	1.82	1.95	2.10	2.25	2.41	2.59	2.77	2.97
7.8	1.70	1.83	1.97	2.12	2.28	2.44	2.62	2.82	3.02	3.24	3.46	3.71
7.9	2.13	2.29	2.46	2.65	2.85	3.06	3.28	3.52	3.77	4.04	4.32	4.62
8.0	2.66	2.87	3.08	3.31	3.56	3.82	4.10	4.39	4.70	5.03	5.38	5.75
8.1	3.33	3.58	3.85	4.14	4.44	4.76	5.10	5.46	5.85	6.25	6.68	7.14
8.2	4.16	4.47	4.80	5.15	5.52	5.92	6.34	6.78	7.25	7.75	8.27	8.82
8.3	5.18	5.56	5.97	6.40	6.86	7.34	7.85	8.39	8.96	9.56	10.2	10.9
8.4	6.43	6.90	7.40	7.93	8.48	9.07	9.69	10.3	11.0	11.7	12.5	13.3
8.5	7.97	8.54	9.14	9.78	10.45	11.16	11.90	12.7	13.5	14.4	15.2	16.2
8.6	9.83	10.5	11.2	12.0	12.8	13.6	14.5	15.5	16.4	17.4	18.5	19.5
8.7	12.07	12.9	13.8	14.7	15.6	16.6	17.6	18.7	19.8	21.0	22.2	23.4
8.8	14.7	15.7	16.7	17.8	18.9	20.0	21.2	22.5	23.7	25.1	26.4	27.8
8.9	17.9	19.0	20.2	21.4	22.7	24.0	25.3	26.7	28.2	29.6	31.1	32.6

[1]   Table provided by Teresa Norberg-King, Duluth, Minnesota. Also see Emerson et al. (1975), Thurston et al. (1974), and USEPA (1985a).

8.9 PRELIMINARY TOXICITY RANGE-FINDING TESTS

8.9.1   USEPA Regional and State personnel generally have observed that it is not necessary to conduct a toxicity range-finding test prior to initiating a static, chronic, definitive toxicity test. However, when preparing to perform a static test with a sample of completely unknown quality, or before initiating a flow-through test, it is advisable to conduct a preliminary toxicity range-finding test.

8.9.2   A toxicity range-finding test ordinarily consists of a down-scaled, abbreviated static acute test in which groups of five organisms are exposed to several widely-spaced sample dilutions in a logarithmic series, such as 100%, 10.0%, 1.00%, and 0.100%, and a control, for 8-24 h. **Caution**: if the sample must also be used for the full-scale definitive test, the 36-h limit on holding time (see Subsection 8.5.4) must not be exceeded before the definitive test is initiated.

8.9.3   It should be noted that the toxicity (LC50) of a sample observed in a range-finding test may be significantly different from the toxicity observed in the follow-up, chronic, definitive test because: (1) the definitive test is longer; and (2) the test may be performed with a sample collected at a different time, and possibly differing significantly in the level of toxicity.

8.10 MULTICONCENTRATION (DEFINITIVE) EFFLUENT TOXICITY TESTS

8.10.1   The tests recommended for use in determining discharge permit compliance in the NPDES program are multiconcentration, or definitive, tests which provide (1) a point estimate of effluent toxicity in terms of an IC25, IC50, or LC50, or (2) a no-observed-effect-concentration (NOEC) defined in terms of mortality, growth, reproduction, and/or teratogenicity and obtained by hypothesis testing. The tests may be static renewal or static non-renewal.

8.10.2   The tests consist of a control and a minimum of five effluent concentrations. USEPA recommends the use of a ≥ 0.5 dilution factor for selecting effluent test concentrations. Effluent test concentrations of 6.25%, 12.5%, 25%, 50%, and 100% are commonly used, however, test concentrations should be selected independently for each test based on the objective of the study, the expected range of toxicity, the receiving water concentration, and any available historical testing information on the effluent. USEPA (2000a) provides additional guidance on choosing appropriate test concentrations.

8.10.3   When these tests are used in determining compliance with permit limits, effluent test concentrations should be selected to bracket the receiving water concentration. This may be achieved by selecting effluent test concentrations in the following manner: (1) 100% effluent, (2) [RWC + 100]/2, (3) RWC, (4) RWC/2, and (5) RWC/4. For example, where the RWC = 50%, appropriate effluent concentrations may be 100%, 75%, 50%, 25%, and 12.5%.

8.10.4   If acute/chronic ratios are to be determined by simultaneous acute and short-term chronic tests with a single species, using the same sample, both types of tests must use the same test conditions, i.e., pH, temperature, water hardness, salinity, etc.

8.11 RECEIVING WATER TESTS

8.11.1   Receiving water toxicity tests generally consist of 100% receiving water and a control. The total salinity of the control should be comparable to the receiving water.

8.11.2   The data from the two treatments are analyzed by hypothesis testing to determine if test organism survival in the receiving water differs significantly from the control. Four replicates and 10 organisms per replicate are required for each treatment (see Summary of Test Conditions and Test Acceptability Criteria in the specific test method).

8.11.3   In cases where the objective of the test is to estimate the degree of toxicity of the receiving water, a definitive, multiconcentration test is performed by preparing dilutions of the receiving water, using a ≥ 0.5 dilution series, with a suitable control water.

SECTION 9

CHRONIC TOXICITY TEST ENDPOINTS AND DATA ANALYSIS

9.1 ENDPOINTS

9.1.1   The objective of chronic aquatic toxicity tests with effluents and pure compounds is to estimate the highest "safe" or "no-effect concentration" of these substances. For practical reasons, the responses observed in these tests are usually limited to hatchability, gross morphological abnormalities, survival, growth, and reproduction, and the results of the tests are usually expressed in terms of the highest toxicant concentration that has no statistically significant observed effect on these responses, when compared to the controls. The terms currently used to define the endpoints employed in the rapid, chronic and sub-chronic toxicity tests have been derived from the terms previously used for full life-cycle tests. As shorter chronic tests were developed, it became common practice to apply the same terminology to the endpoints. The terms used in this manual are as follows:

9.1.1.1   Safe Concentration - The highest concentration of toxicant that will permit normal propagation of fish and other aquatic life in receiving waters. The concept of a "safe concentration" is a biological concept, whereas the "no-observed-effect concentration" (below) is a statistically defined concentration.

9.1.1.2   No-Observed-Effect-Concentration (NOEC) - The highest concentration of toxicant to which organisms are exposed in a full life-cycle or partial life-cycle (short-term) test, that causes no observable adverse effects on the test organisms (i.e., the highest concentration of toxicant in which the values for the observed responses are not statistically significantly different from the controls). This value is used, along with other factors, to determine toxicity limits in permits.

9.1.1.3   Lowest-Observed-Effect-Concentration (LOEC) - The lowest concentration of toxicant to which organisms are exposed in a life-cycle or partial life-cycle (short-term) test, which causes adverse effects on the test organisms (i.e., where the values for the observed responses are statistically significantly different from the controls).

9.1.1.4   Effective Concentration (EC) - A point estimate of the toxicant concentration that would cause an observable adverse affect on a quantal, "all or nothing," response (such as death, immobilization, or serious incapacitation) in a given percent of the test organisms, calculated by point estimation techniques. If the observable effect is death or immobility, the term, Lethal Concentration (LC), should be used (see Subsection 9.1.1.5). A certain EC or LC value might be judged from a biological standpoint to represent a threshold concentration, or lowest concentration that would cause an adverse effect on the observed response.

9.1.1.5   Lethal Concentration (LC) - The toxicant concentration that would cause death in a given percent of the test population. Identical to EC when the observable adverse effect is death. For example, the LC50 is the concentration of toxicant that would cause death in 50% of the test population.

9.1.1.6   Inhibition Concentration (IC) - The toxicant concentration that would cause a given percent reduction in a nonquantal biological measurement for the test population. For example, the IC25 is the concentration of toxicant that would cause a 25% reduction in mean young per female or in growth for the test population, and the IC50 is the concentration of toxicant that would cause a 50% reduction in the mean population responses.

9.2 RELATIONSHIP BETWEEN ENDPOINTS DETERMINED BY HYPOTHESIS TESTING AND POINT ESTIMATION TECHNIQUES

9.2.1   If the objective of chronic aquatic toxicity tests with effluents and pure compounds is to estimate the highest "safe or no-effect concentration" of these substances, it is imperative to understand how the statistical endpoints of these tests are related to the "safe" or "no-effect" concentration. NOECs and LOECs are determined by hypothesis testing (Dunnett's Test, a t test with the Bonferroni adjustment, Steel's Many-One Rank Test, or the Wilcoxon Rank

Sum Test with Bonferroni adjustment), whereas LCs, ICs, and ECs are determined by point estimation techniques (Probit Analysis, the Spearman-Karber Method, the Trimmed Spearman-Karber Method, the Graphical Method or Linear Interpolation Method). There are inherent differences between the use of a NOEC or LOEC derived from hypothesis testing to estimate a "safe" concentration, and the use of a LC, IC, EC, or other point estimates derived from curve fitting, interpolation, etc.

9.2.2   Most point estimates, such as the LC, IC, or EC are derived from a mathematical model that assumes a continuous dose-response relationship. By definition, any LC, IC, or EC value is an estimate of some amount of adverse effect. Thus the assessment of a "safe" concentration must be made from a biological standpoint rather than with a statistical test. In this instance, the biologist must determine some amount of adverse effect that is deemed to be "safe," in the sense that from a practical biological viewpoint it will not affect the normal propagation of fish and other aquatic life in receiving waters.

9.2.3   The use of NOECs and LOECs, on the other hand, assumes either (1) a continuous dose-response relationship, or (2) a non-continuous (threshold) model of the dose-response relationship.

9.2.3.1   In the case of a continuous dose-response relationship, it is also assumed that adverse effects that are not "statistically observable" are also not important from a biological standpoint, since they are not pronounced enough to test as statistically significant against some measure of the natural variability of the responses.

9.2.3.2   In the case of non-continuous dose-response relationships, it is assumed that there exists a true threshold, or concentration below which there is no adverse effect on aquatic life, and above which there is an adverse effect. The purpose of the statistical analysis in this case is to estimate as closely as possible where that threshold lies.

9.2.3.3   In either case, it is important to realize that the amount of adverse effect that is statistically observable (LOEC) or not observable (NOEC) is highly dependent on all aspects of the experimental design, such as the number of concentrations of toxicant, number of replicates per concentration, number of organisms per replicate, and use of randomization. Other factors that affect the sensitivity of the test include the choice of statistical analysis, the choice of an alpha level, and the amount of variability between responses at a given concentration.

9.2.3.4   Where the assumption of a continuous dose-response relationship is made, by definition some amount of adverse effect might be present at the NOEC, but is not great enough to be detected by hypothesis testing.

9.2.3.5   Where the assumption of a noncontinuous dose-response relationship is made, the NOEC would indeed be an estimate of a "safe" or "no-effect" concentration if the amount of adverse effect that appears at the threshold is great enough to test as statistically significantly different from the controls in the face of all aspects of the experimental design mentioned above. If, however, the amount of adverse effect at the threshold were not great enough to test as statistically different, some amount of adverse effect might be present at the NOEC. In any case, the estimate of the NOEC with hypothesis testing is always dependent on the aspects of the experimental design mentioned above. For this reason, the reporting and examination of some measure of the sensitivity of the test (either the minimum significant difference or the percent change from the control that this minimum difference represents) is extremely important.

9.2.4   In summary, the assessment of a "safe" or "no-effect" concentration cannot be made from the results of statistical analysis alone, unless (1) the assumptions of a strict threshold model are accepted, and (2) it is assumed that the amount of adverse effect present at the threshold is statistically detectable by hypothesis testing. In this case, estimates obtained from a statistical analysis are indeed estimates of a "no-effect" concentration. If the assumptions are not deemed tenable, then estimates from a statistical analysis can only be used in conjunction with an assessment from a biological standpoint of what magnitude of adverse effect constitutes a "safe" concentration. In this instance, a "safe" concentration is not necessarily a truly "no-effect" concentration, but rather a concentration at which the effects are judged to be of no biological significance.

9.2.5   A better understanding of the relationship between endpoints derived by hypothesis testing (NOECs) and point estimation techniques (LCs, ICs, and ECs) would be very helpful in choosing methods of data analysis. Norberg-King (1991) reported that the IC25s were comparable to the NOECs for 23 effluent and reference toxicant data sets analyzed. The data sets included short-term chronic toxicity tests for the sea urchin, *Arbacia punctulata*, the sheepshead minnow, *Cyprinodon variegatus*, and the red macroalga, *Champia parvula*. Birge et al. (1985) reported that LC1s derived from Probit Analyses of data from short-term embryo-larval tests with reference toxicants were comparable to NOECs for several organisms. Similarly, USEPA (1988d) reported that the IC25s were comparable to the NOECs for a set of daphnia, *Ceriodaphnia dubia* chronic tests with a single reference toxicant. However, the scope of these comparisons was very limited, and sufficient information is not yet available to establish an overall relationship between these two types of endpoints, especially when derived from effluent toxicity test data.

9.3   PRECISION

9.3.1   HYPOTHESIS TESTS

9.3.1.1   When hypothesis tests are used to analyze toxicity test data, it is not possible to express precision in terms of a commonly used statistic. The results of the test are given in terms of two endpoints, the No-Observed-Effect Concentration (NOEC) and the Lowest-Observed-Effect Concentration (LOEC). The NOEC and LOEC are limited to the concentrations selected for the test. The width of the NOEC-LOEC interval is a function of the dilution series, and differs greatly depending on whether a dilution factor of 0.3 or 0.5 is used in the test design. Therefore, **USEPA recommends the use of the ≥ 0.5 dilution factor** (see Section 4, Quality Assurance). It is not possible to place confidence limits on the NOEC and LOEC derived from a given test, and it is difficult to quantify the precision of the NOEC-LOEC endpoints between tests. If the data from a series of tests performed with the same toxicant, toxicant concentrations, and test species, were analyzed with hypothesis tests, precision could only be assessed by a qualitative comparison of the NOEC-LOEC intervals, with the understanding that maximum precision would be attained if all tests yielded the same NOEC-LOEC interval. In practice, the precision of results of repetitive chronic tests is considered acceptable if the NOECs vary by no more than one concentration interval above or below a central tendency. Using these guidelines, the "normal" range of NOECs from toxicity tests using a 0.5 dilution factor (two-fold difference between adjacent concentrations), would be four-fold.

9.3.2   POINT ESTIMATION TECHNIQUES

9.3.2.1   Point estimation techniques have the advantage of providing a point estimate of the toxicant concentration causing a given amount of adverse (inhibiting) effect, the precision of which can be quantitatively assessed (1) within tests by calculation of 95% confidence limits, and (2) across tests by calculating a standard deviation and coefficient of variation.

9.3.2.2   It should be noted that software used to calculate point estimates occasionally may not provide associated 95% confidence intervals. This situation may arise when test data do not meet specific assumptions required by the statistical methods, when point estimates are outside of the test concentration range, and when specific limitations imposed by the software are encountered. USEPA (2000a) provides guidance on confidence intervals under these circumstances.

9.4   DATA ANALYSIS

9.4.1   ROLE OF THE STATISTICIAN

9.4.1.1   The use of the statistical methods described in this manual for routine data analysis does not require the assistance of a statistician. However, the interpretation of the results of the analysis of the data from any of the toxicity tests described in this manual can become problematic because of the inherent variability and sometimes unavoidable anomalies in biological data. If the data appear unusual in any way, or fail to meet the necessary

assumptions, a statistician should be consulted. Analysts who are not proficient in statistics are strongly advised to seek the assistance of a statistician before selecting the method of analysis and using any of the results.

9.4.1.2   The statistical methods recommended in this manual are not the only possible methods of statistical analysis. Many other methods have been proposed and considered. Certainly there are other reasonable and defensible methods of statistical analysis for this kind of toxicity data. Among alternative hypothesis tests some, like Williams' Test, require additional assumptions, while others, like the bootstrap methods, require computer-intensive computations. Alternative point estimations approaches most probably would require the services of a statistician to determine the appropriateness of the model (goodness of fit), higher order linear or nonlinear models, confidence intervals for estimates generated by inverse regression, etc. In addition, point estimation or regression approaches would require the specification by biologists or toxicologists of some low level of adverse effect that would be deemed acceptable or safe. The statistical methods contained in this manual have been chosen because they are (1) applicable to most of the different toxicity test data sets for which they are recommended, (2) powerful statistical tests, (3) hopefully "easily" understood by nonstatisticians, and (4) amenable to use without a computer, if necessary.

9.4.2   PLOTTING THE DATA

9.4.2.1   The data should be plotted, both as a preliminary step to help detect problems and unsuspected trends or patterns in the responses, and as an aid in interpretation of the results. Further discussion and plotted sets of data are included in the methods and the Appendices.

9.4.3   DATA TRANSFORMATIONS

9.4.3.1   Transformations of the data, (e.g., arc sine square root and logs), are used where necessary to meet assumptions of the proposed analyses, such as the requirement for normally distributed data.

9.4.4   INDEPENDENCE, RANDOMIZATION, AND OUTLIERS

9.4.4.1   Statistical independence among observations is a critical assumption in all statistical analysis of toxicity data. One of the best ways to ensure independence is to properly follow rigorous randomization procedures. Randomization techniques should be employed at the start of the test, including the randomization of the placement of test organisms in the test chambers and randomization of the test chamber location within the array of chambers. Discussions of statistical independence, outliers and randomization, and a sample randomization scheme, are included in Appendix A.

9.4.5   REPLICATION AND SENSITIVITY

9.4.5.1   The number of replicates employed for each toxicant concentration is an important factor in determining the sensitivity of chronic toxicity tests. Test sensitivity generally increases as the number of replicates is increased, but the point of diminishing returns in sensitivity may be reached rather quickly. The level of sensitivity required by a hypothesis test or the confidence interval for a point estimate will determine the number of replicates, and should be based on the objectives for obtaining the toxicity data.

9.4.5.2   In a statistical analysis of toxicity data, the choice of a particular analysis and the ability to detect departures from the assumptions of the analysis, such as the normal distribution of the data and homogeneity of variance, is also dependent on the number of replicates. More than the minimum number of replicates may be required in situations where it is imperative to obtain optimal statistical results, such as with tests used in enforcement cases or when it is not possible to repeat the tests. For example, when the data are analyzed by hypothesis testing, the nonparametric alternatives cannot be used unless there are at least four replicates at each toxicant concentration.

9.4.6 RECOMMENDED ALPHA LEVELS

9.4.6.1   The data analysis examples included in the manual specify an alpha level of 0.01 for testing the assumptions of hypothesis tests and an alpha level of 0.05 for the hypothesis tests themselves. These levels are common and well accepted levels for this type of analysis and are presented as a recommended minimum significance level for toxicity data analysis.

9.5   CHOICE OF ANALYSIS

9.5.1   The recommended statistical analysis of most data from chronic toxicity tests with aquatic organisms follows a decision process illustrated in the flowchart in Figure 2. An initial decision is made to use point estimation techniques (the Probit Analysis, the Spearman-Karber Method, the Trimmed Spearman-Karber Method, the Graphical Method, or Linear Interpolation Method) and/or to use hypothesis testing (Dunnett's Test, the t test with the Bonferroni adjustment, Steel's Many-one Rank Test, or Wilcoxon Rank Sum Test with the Bonferroni adjustment). **NOTE: For the NPDES Permit Program, the point estimation techniques are the preferred statistical methods in calculating end points for effluent toxicity tests.** If hypothesis testing is chosen, subsequent decisions are made on the appropriate procedure for a given set of data, depending on the results of tests of assumptions, as illustrated in the flowchart. A specific flow chart is included in the analysis section for each test.

9.5.2   Since a single chronic toxicity test might yield information on more than one parameter (such as survival, growth, and reproduction), the lowest estimate of a "no-observed-effect concentration" for any of the responses would be used as the "no observed effect concentration" for each test. It follows logically that in the statistical analysis of the data, concentrations that had a significant toxic effect on one of the observed responses would not be subsequently tested for an effect on some other response. This is one reason for excluding concentrations that have shown a statistically significant reduction in survival from a subsequent hypothesis test for effects on another parameter such as reproduction. A second reason is that the exclusion of such concentrations usually results in a more powerful and appropriate statistical analysis. In performing the point estimation techniques recommended in this manual, an all-data approach is used. For example, data from concentrations above the NOEC for survival are included in determining ICp estimates using the Linear Interpolation Method.

9.5.3   ANALYSIS OF GROWTH AND REPRODUCTION DATA

9.5.3.1   Growth data from the sheepshead minnow, *Cyprinodon variegatus*, and inland silverside, *Menidia beryllina*, larval survival and growth tests, and the mysid, *Mysidopsis bahia*, survival, growth, and fecundity test, are analyzed using hypothesis testing according to the flowchart in Figure 2. The above mentioned growth data may also be analyzed by generating a point estimate with the Linear Interpolation Method. Data from effluent concentrations that have tested significantly different from the control for survival are excluded from further hypothesis tests concerning growth effects. Growth is defined as the change in dry weight of the orginal number of test organisms when group weights are obtained. When analyzing the data using point estimating techniques, data from all concentrations are included in the analysis.

9.5.3.2   Fecundity data from the mysid, *Mysidopsis bahia*, test may be analyzed using hypothesis testing after an arc sine transformation according to the flowchart in Figure 2. The fecundity data from the mysid test may also be analyzed by generating a point estimate with the Linear Interpolation Method.

9.5.3.3   Reproduction data from the red macroalga, *Champia parvula*, test are analyzed using hypothesis testing as illustrated in Figure 2. The reproduction data from the red macroalga test may also be analyzed by generating a point estimate with the Linear Interpolation Method.

DATA (SURVIVAL, GROWTH, REPRODUCTION, ETC.)

POINT ESTIMATION

ENDPOINT ESTIMATE
LC, EC, IC

HYPOTHESIS TESTING

TRANSFORMATION?

SHAPIRO-WILK'S TEST → NON-NORMAL DISTRIBUTION

NORMAL DISTRIBUTION

BARTLETT'S TEST → HETEROGENEOUS VARIANCE

HOMOGENEOUS VARIANCE

NO STATISTICAL ANALYSIS RECOMMENDED ← NO — 4 OR MORE REPLICATES?

YES

EQUAL NUMBER OF REPLICATES?

EQUAL NUMBER OF REPLICATES?

NO   YES   YES   NO

T-TEST WITH BONFERRONI ADJUSTMENT

DUNNETT'S TEST

STEEL'S MANY-ONE RANK TEST

WILCOXON RANK SUM TEST WITH BONFERRONI ADJUSTMENT
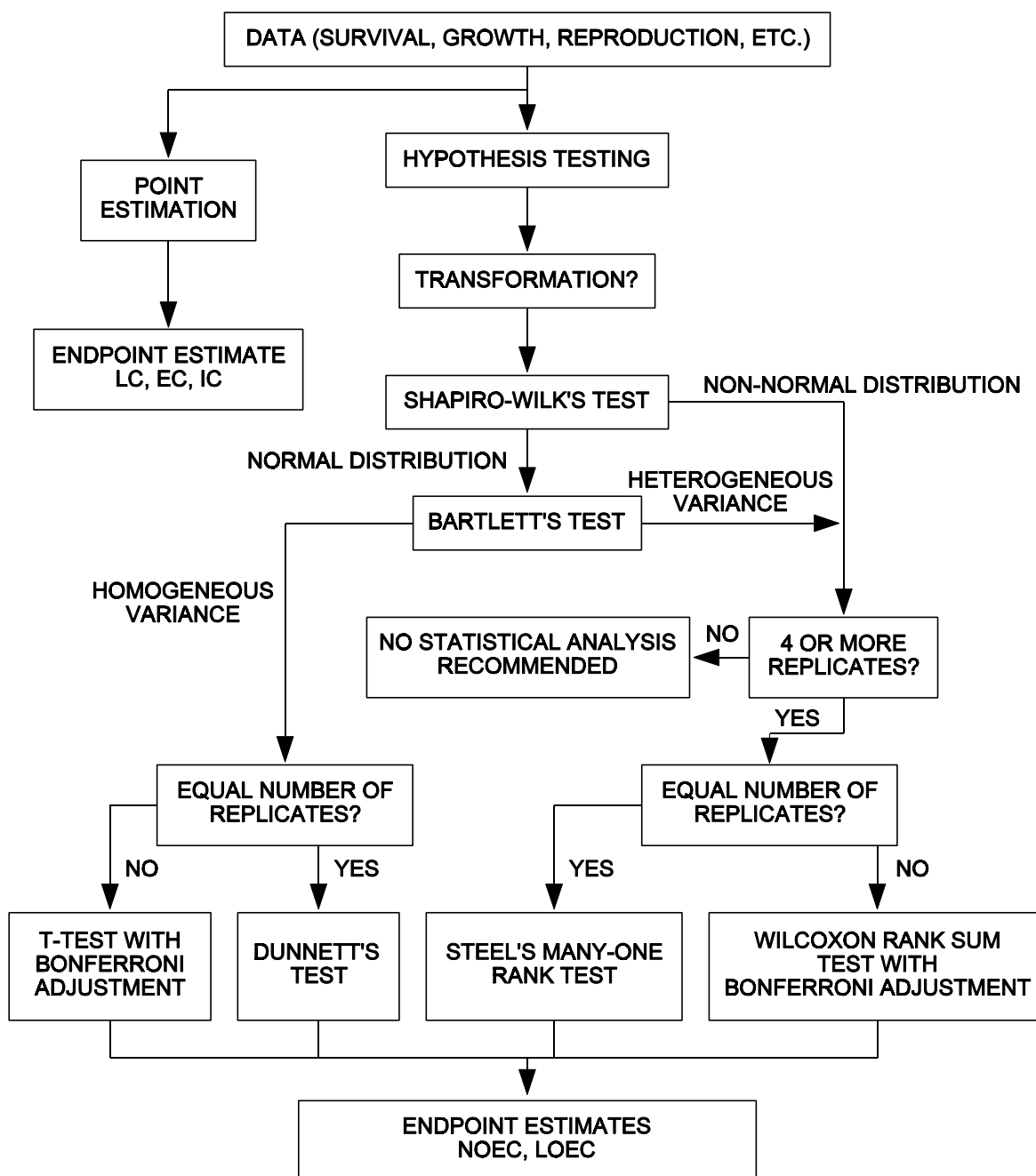
ENDPOINT ESTIMATES
NOEC, LOEC

Figure 2. Flowchart for statistical analysis of test data

9.5.4   ANALYSIS OF THE SEA URCHIN, *ARBACIA PUNCTULATA*, FERTILIZATION DATA

9.5.4.1   Data from the sea urchin, *Arbacia punctulata*, fertilization test may be analyzed by hypothesis testing after an arc sine transformation according to the flowchart in Figure 2. The fertilization data from the sea urchin test may also be analyzed by generating a point estimate with the Linear Interpolation Method.

9.5.5   ANALYSIS OF MORTALITY DATA

9.5.5.1   Mortality data are analyzed by Probit Analysis, if appropriate, or other point estimation techniques, (i.e., the Spearman-Karber Method, the Trimmed Spearman-Karber Method, or the Graphical Method) (see Appendices H-K) (see discussion below). The mortality data can also be analyzed by hypothesis testing, after an arc sine square root transformation (see Appendices B-F), according to the flowchart in Figure 2.

9.6   HYPOTHESIS TESTS

9.6.1   DUNNETT'S PROCEDURE

9.6.1.1   Dunnett's Procedure is used to determine the NOEC. The procedure consists of an analysis of variance (ANOVA) to determine the error term, which is then used in a multiple comparison procedure for comparing each of the treatment means with the control mean, in a series of paired tests (see Appendix C). Use of Dunnett's Procedure requires at least three replicates per treatment to check the assumptions of the test. In cases where the numbers of data points (replicates) for each concentration are not equal, a t test may be performed with Bonferroni's adjustment for multiple comparisons (see Appendix D), instead of using Dunnett's Procedure.

9.6.1.2   The assumptions upon which the use of Dunnett's Procedure is contingent are that the observations within treatments are normally distributed, with homogeneity of variance. Before analyzing the data, these assumptions must be tested using the procedures provided in Appendix B.

9.6.1.3   If, after suitable transformations have been carried out, the normality assumptions have not been met, Steel's Many-one Rank Test should be used if there are four or more data points (replicates) per toxicant concentration. If the numbers of data points for each toxicant concentration are not equal, the Wilcoxon Rank Sum Test with Bonferroni's adjustment should be used (see Appendix F).

9.6.1.4   Some indication of the sensitivity of the analysis should be provided by calculating (1) the minimum difference between means that can be detected as statistically significant, and (2) the percent change from the control mean that this minimum difference represents for a given test.

9.6.1.5   A step-by-step example of the use of Dunnett's Procedure is provided in Appendix C.

9.6.2   T TEST WITH THE BONFERRONI ADJUSTMENT

9.6.2.1   The t test with the Bonferroni adjustment is used as an alternative to Dunnett's Procedure when the number of replicates is not the same for all concentrations. This test sets an upper bound of alpha on the overall error rate, in contrast to Dunnett's Procedure, for which the overall error rate is fixed at alpha. Thus, Dunnett's Procedure is a more powerful test.

9.6.2.2   The assumptions upon which the use of the t test with the Bonferroni adjustment is contingent are that the observations within treatments are normally distributed, with homogeneity of variance. These assumptions must be tested using the procedures provided in Appendix B.

9.6.2.3   The estimate of the safe concentration derived from this test is reported in terms of the NOEC. A step-by-step example of the use of a t-test with the Bonferroni adjustment is provided in Appendix D.

9.6.3   STEEL'S MANY-ONE RANK TEST

9.6.3.1   Steel's Many-one Rank Test is a multiple comparison procedure for comparing several treatments with a control. This method is similar to Dunnett's procedure, except that it is not necessary to meet the assumption of normality. The data are ranked, and the analysis is performed on the ranks rather than on the data themselves. If the data are normally or nearly normally distributed, Dunnett's Procedure would be more sensitive (would detect smaller differences between the treatments and control). For data that are not normally distributed, Steel's Many-one Rank Test can be much more efficient (Hodges and Lehmann, 1956).

9.6.3.2   It is necessary to have at least four replicates per toxicant concentration to use Steel's test. Unlike Dunnett's procedure, the sensitivity of this test cannot be stated in terms of the minimum difference between treatment means and the control mean that can be detected as statistically significant.

9.6.3.3   The estimate of the safe concentration is reported as the NOEC. A step-by-step example of the use of Steel's Many-One Rank Test is provided in Appendix E.

9.6.4   WILCOXON RANK SUM TEST WITH THE BONFERRONI ADJUSTMENT

9.6.4.1   The Wilcoxon Rank Sum Test is a nonparametric test for comparing a treatment with a control. The data are ranked and the analysis proceeds exactly as in Steel's Test except that Bonferroni's adjustment for multiple comparisons is used instead of Steel's tables. When Steel's test can be used (i.e., when there are equal numbers of data points per toxicant concentration), it will be more powerful (able to detect smaller differences as statistically significant) than the Wilcoxon Rank Sum Test with Bonferroni's adjustment.

9.6.4.2   The estimate of the safe concentration is reported as the NOEC. A step-by-step example of the use of the Wilcoxon Rank Sum Test with Bonferroni adjustment is provided in Appendix F.

9.6.5   A CAUTION IN THE USE OF HYPOTHESIS TESTING

9.6.5.1   If in the calculation of an NOEC by hypothesis testing, two tested concentrations cause statistically significant adverse effects, but an intermediate concentration did not cause statistically significant effects, the results should be used with extreme caution.

9.7   POINT ESTIMATION TECHNIQUES

9.7.1   PROBIT ANALYSIS

9.7.1.1   Probit Analysis is used to estimate an LC1, LC50, EC1, or EC50 and the associated 95% confidence interval. The analysis consists of adjusting the data for mortality in the control, and then using a maximum likelihood technique to estimate the parameters of the underlying log tolerance distribution, which is assumed to have a particular shape.

9.7.1.2   The assumption upon which the use of Probit Analysis is contingent is a normal distribution of log tolerances. If the normality assumption is not met, and at least two partial mortalities are not obtained, Probit Analysis should not be used. It is important to check the results of Probit Analysis to determine if use of the analysis is appropriate. The chi-square test for heterogeneity provides a good test of appropriateness of the analysis. The computer program (see discussion, Appendix H) checks the chi-square statistic calculated for the data set against the tabular value, and provides an error message if the calculated value exceeds the tabular value.

9.7.1.3   A discussion of Probit Analysis, and examples of computer program input and output, are found in Appendix H.

9.7.1.4   In cases where Probit Analysis is not appropriate, the LC50 and confidence interval may be estimated by the Spearman-Karber Method (Appendix I) or the Trimmed Spearman-Karber Method (Appendix J). If a test results in 100% survival and 100% mortality in adjacent treatments (all or nothing effect), the LC50 may be estimated using the Graphical Method (Appendix K).

9.7.2   LINEAR INTERPOLATION METHOD

9.7.2.1   The Linear Interpolation Method (see Appendix L) is a procedure to calculate a point estimate of the effluent or other toxicant concentration [Inhibition Concentration, (IC)] that causes a given percent reduction (e.g., 25%, 50%, etc.) in the reproduction, growth, fertilization, or fecundity of the test organisms. The procedure was designed for general applicability in the analysis of data from short-term chronic toxicity tests.

9.7.2.2   Use of the Linear Interpolation Method is based on the assumptions that the responses (1) are monotonically non-increasing (the mean response for each higher concentration is less than or equal to the mean response for the previous concentration), (2) follow a piece-wise linear response function, and (3) are from a random, independent, and representative sample of test data. The assumption for piece-wise linear response cannot be tested statistically, and no defined statistical procedure is provided to test the assumption for monotonicity. Where the observed means are not strictly monotonic by examination, they are adjusted by smoothing. In cases where the responses at the low toxicant concentrations are much higher than in the controls, the smoothing process may result in a large upward adjustment in the control mean.

9.7.2.3   The inability to test the monotonicity and piece wise linear response assumptions for this method makes it difficult to assess when the method is, or is not, producing reliable results. Therefore, the method should be used with caution when the results of a toxicity test approach an "all or nothing" response from one concentration to the next in the concentration series, and when it appears that there is a large deviation from monotonicity. See Appendix L for a more detailed discussion of the use of this method and a computer program available for performing calculations.

SECTION 10

REPORT PREPARATION AND TEST REVIEW

10.1 REPORT PREPARATION

The toxicity data are reported, together with other appropriate data. The following general format and content are recommended for the report:

10.1.1 INTRODUCTION

    1.   Permit number
    2.   Toxicity testing requirements of permit
    3.   Plant location
    4.   Name of receiving water body
    5.   Contract Laboratory (if the test was performed under contract)
            a.   Name of firm
            b.   Phone number
            c.   Address
    6.   Objective of test

10.1.2 PLANT OPERATIONS

    1.   Product(s)
    2.   Raw materials
    3.   Operating schedule
    4.   Description of waste treatment
    5.   Schematic of waste treatment
    6.   Retention time (if applicable)
    7.   Volume of waste flow (MGD, CFS, GPM)
    8.   Design flow of treatment facility at time of sampling

10.1.3 SOURCE OF EFFLUENT, RECEIVING WATER, AND DILUTION WATER

    1.   Effluent Samples
            a.   Sampling point (including latitude and longitude)
            b.   Collection dates and times
            c.   Sample collection method
            d.   Physical and chemical data
            e.   Mean daily discharge on sample collection date
            f.   Lapsed time from sample collection to delivery
            g.   Sample temperature when received at the laboratory

    2.   Receiving Water Samples
            a.   Sampling point (including latitude and longitude)
             b.   Collection dates and times
            c.   Sample collection method
            d.   Physical and chemical data
            e.   Tide stages
            f.   Sample temperature when received at the laboratory
            g.   Lapsed time from sample collection to delivery

3.   Dilution Water Samples
      a.   Source
      b.   Collection date and time
      c.   Pretreatment
      d.   Physical and chemical characteristics

10.1.4 TEST METHODS

1.   Toxicity test method used (title, number, source)
2.   Endpoint(s) of test
3.   Deviation(s) from reference method, if any, and the reason(s)
4.   Date and time test started
5.   Date and time test terminated
6.   Type of volume and test chambers
7.   Volume of solution used per chamber
8.   Number of organisms used per test chamber
9.   Number of replicate test chambers per treatment
10.   Acclimation of test organisms (temperature and salinity mean and range)
11.   Test temperature (mean and range)
12.   Specify if aeration was needed
13.   Feeding frequency, and amount and type of food
14.   Test salinity (mean and range)
15.   Specify if (and how) pH control measures were implemented

10.1.5 TEST ORGANISMS

1.   Scientific name and how determined
2.   Age
3.   Life stage
4.   Mean length and weight (where applicable)
5.   Source
6.   Diseases and treatment (where applicable)
7.   Taxonomic key used for species identification

10.1.6 QUALITY ASSURANCE

1.   Reference toxicant used routinely; source
2.   Date and time of most recent reference toxicant test; test results and current control (cusum) chart
3.   Dilution water used in reference toxicant test
4.   Results (NOEC or, where applicable, LOEC, LC50, EC50, IC25 and/or IC50); report percent minimum significant difference (PMSD) calculated for sublethal endpoints determined by hypothesis testing in reference toxicant test
5.   Physical and chemical methods used

10.1.7 RESULTS

1.   Provide raw toxicity data in tabular form, including daily records of affected organisms in each concentration (including controls) and replicate, and in graphical form (plots of toxicity data)
2.   Provide table of LC50s, NOECs, IC25, IC50, etc. (as required in the applicable NPDES permit)
3.   Indicate statistical methods to calculate endpoints
4.   Provide summary table of physical and chemical data
5.   Tabulate QA data
6.   Provide percent minimum significant difference (PMSD) calculated for sublethal endpoints

10.1.8 CONCLUSIONS AND RECOMMENDATIONS

    1.   Relationship between test endpoints and permit limits.
    2.   Action to be taken.

10.2 TEST REVIEW

10.2.1 Test review is an important part of an overall quality assurance program (Section 4) and is necessary for ensuring that all test results are reported accurately. Test review should be conducted on each test by both the testing laboratory and the regulatory authority.

10.2.2 SAMPLING AND HANDLING

10.2.2.1 The collection and handling of samples are reviewed to verify that the sampling and handling procedures given in Section 8 were followed. Chain-of-custody forms are reviewed to verify that samples were tested within allowable sample holding times (Subsection 8.5.4). Any deviations from the procedures given in Section 8 should be documented and described in the data report (Subsection 10.1).

10.2.3 TEST ACCEPTABILITY CRITERIA

10.2.3.1 Test data are reviewed to verify that test acceptability criteria (TAC) requirements for a valid test have been met. Any test not meeting the minimum test acceptability criteria is considered invalid. All invalid tests must be repeated with a newly collected sample.

10.2.4 TEST CONDITIONS

10.2.4.1 Test conditions are reviewed and compared to the specifications listed in the summary of test condition tables provided for each method. Physical and chemical measurements taken during the test (e.g., temperature, pH, and DO) also are reviewed and compared to specified ranges. Any deviations from specifications should be documented and described in the data report (Subsection 10.1).

10.2.4.2 The summary of test condition tables presented for each method identify test conditions as required or recommended. For WET test data submitted under NPDES permits, all required test conditions must be met or the test is considered invalid and must be repeated with a newly collected sample.   Deviations from recommended test conditions must be evaluated on a case-by-case basis to determine the validity of test results. Deviations from recommended test conditions may or may not invalidate a test result depending on the degree of the departure and the objective of the test. The reviewer should consider the degree of the deviation and the potential or observed impact of the deviation on the test result before rejecting or accepting a test result as valid. For example, if dissolved oxygen is measured below 4.0 mg/L in one test chamber, the reviewer should consider whether any observed mortality in that test chamber corresponded with the drop in dissolved oxygen.

10.2.4.3 Whereas slight deviations in test conditions may not invalidate an individual test result, test condition deviations that continue to occur frequently in a given laboratory may indicate the need for improved quality control in that laboratory.

10.2.5 STATISTICAL METHODS

10.2.5.1 The statistical methods used for analyzing test data are reviewed to verify that the recommended flowcharts for statistical analysis were followed. Any deviation from the recommended flowcharts for selection of statistical methods should be noted in the data report. Statistical methods other than those recommended in the statistical flowcharts may be appropriate (see Subsection 9.4.1.2), however, the laboratory must document the use of and provide the rationale for the use of any alternate statistical method. In all cases (flowchart recommended

methods or alternate methods), reviewers should verify that the necessary assumptions are met for the statistical method used.

10.2.6 CONCENTRATION-RESPONSE RELATIONSHIPS

10.2.6.1 The concept of a concentration-response, or more classically, a dose-response relationship is "the most fundamental and pervasive one in toxicology" (Casarett and Doull, 1975). This concept assumes that there is a causal relationship between the dose of a toxicant (or concentration for toxicants in solution) and a measured response. A response may be any measurable biochemical or biological parameter that is correlated with exposure to the toxicant. The classical concentration-response relationship is depicted as a sigmoidal shaped curve, however, the particular shape of the concentration-response curve may differ for each coupled toxicant and response pair. In general, more severe responses (such as acute effects) occur at higher concentrations of the toxicant, and less severe responses (such as chronic effects) occur at lower concentrations. A single toxicant also may produce multiple responses, each characterized by a concentration-response relationship. A corollary of the concentration-response concept is that every toxicant should exhibit a concentration-response relationship, given that the appropriate response is measured and given that the concentration range evaluated is appropriate. Use of this concept can be helpful in determining whether an effluent possesses toxicity and in identifying anomalous test results.

10.2.6.2 The concentration-response relationship generated for each multi-concentration test must be reviewed to ensure that calculated test results are interpreted appropriately. USEPA (2000a) provides guidance on evaluating concentration-response relationships to assist in determining the validity of WET test results. All WET test results (from multi-concentration tests) reported under the NPDES program should be reviewed and reported according to USEPA guidance on the evaluation of concentration-response relationships (USEPA, 2000a). This guidance provides review steps for 10 different concentration-response patterns that may be encountered in WET test data. Based on the review, the guidance provides one of three determinations: that calculated effect concentrations are reliable and should be reported, that calculated effect concentrations are anomalous and should be explained, or that the test was inconclusive and the test should be repeated with a newly collected sample. It should be noted that the determination of a valid concentration-response relationship is not always clear cut. Data from some tests may suggest consultation with professional toxicologists and/or regulatory officials. Tests that exhibit unexpected concentration-response relationships also may indicate a need for further investigation and possible retesting.

10.2.7 REFERENCE TOXICANT TESTING

10.2.7.1 Test review of a given effluent or receiving water test should include review of the associated reference toxicant test and current control chart. Reference toxicant testing and control charting is required for documenting the quality of test organisms (Subsection 4.7) and ongoing laboratory performance (Subsection 4.16). The reviewer should verify that a quality control reference toxicant test was conducted according to the specified frequency required by the permitting authority or recommended by the method (e.g., monthly). The test acceptability criteria, test conditions, concentration-response relationship, and test sensitivity of the reference toxicant test are reviewed to verify that the reference toxicant test conducted was a valid test. The results of the reference toxicant test are then plotted on a control chart (see Subsection 4.16) and compared to the current control chart limits (± 2 standard deviations).

10.2.7.2 Reference toxicant tests that fall outside of recommended control chart limits are evaluated to determine the validity of associated effluent and receiving water tests (see Subsection 4.16). An out of control reference toxicant test result does not necessarily invalidate associated test results. The reviewer should consider the degree to which the reference toxicant test result fell outside of control chart limits, the width of the limits, the direction of the deviation (toward increasing test organism sensitivity or toward decreasing test organism sensitivity), the test conditions of both the effluent test and the reference toxicant test, and the objective of the test. More frequent and/or concurrent reference toxicant testing may be advantageous if recent problems (e.g., invalid tests, reference toxicant test results outside of control chart limits, reduced health of organism cultures, or increased within-test variability) have been identified in testing.

10.2.8   TEST VARIABILITY

10.2.8.1   The within-test variability of individual tests should be reviewed. Excessive within-test variability may invalidate a test result and warrant retesting. For evaluating within-test variability, reviewers should consult EPA guidance on upper and lower percent minimum significant difference (PMSD) bounds (USEPA, 2000b).

10.2.8.2   When NPDES permits require sublethal hypothesis testing endpoints from Methods 1006.0 or 1007.0 (e.g., growth NOECs and LOECs), within-test variability must be reviewed and variability criteria must be applied as described in this section (10.2.8.2). When the methods are used for non-regulatory purposes, the variability criteria herein are recommended but are not required, and their use (or the use of alternative variability criteria) may depend upon the intended uses of the test results and the requirements of any applicable data quality objectives and quality assurance plan.

10.2.8.2.1   To measure test variability, calculate the percent minimum significant difference (PMSD) achieved in the test. The PMSD is the smallest percentage decrease in growth or reproduction from the control that could be determined as statistically significant in the test. The PMSD is calculated as 100 times the minimum significant difference (MSD) divided by the control mean. The equation and examples of MSD calculations are shown in Appendix C. PMSD may be calculated legitimately as a descriptive statistic for within-test variability, even when the hypothesis test is conducted using a non-parametric method. The PMSD bounds were based on a representative set of tests, including tests for which a non-parametric method was required for determining the NOEC or LOEC. The conduct of hypothesis testing to determine test results should follow the statistical flow charts provided for each method. That is, when test data fail to meet assumptions of normality or heterogeneity of variance, a non-parametric method (determined following the statistical flowchart for the method) should be used to calculate test results, but the PMSD may be calculated as described above (using parametric methods) to provide a measure of test variability.

10.2.8.2.2   Compare the PMSD measured in the test with the upper PMSD bound variability criterion listed in Table 6. When the test PMSD exceeds the upper bound, the variability among replicates is unusually large for the test method. Such a test should be considered insufficiently sensitive to detect toxic effects on growth or reproduction of substantial magnitude. A finding of toxicity at a particular concentration may be regarded as trustworthy, but a finding of "no toxicity" or "no statistically significant toxicity" at a particular concentration should not be regarded as a reliable indication that there is no substantial toxic effect on growth or reproduction at that concentration.

10.2.8.2.3   If the PMSD measured for the test is less than or equal to the upper PMSD bound variability criterion in Table 6, then the test's variability measure lies within normal bounds and the effect concentration estimate (e.g., NOEC or LOEC) would normally be accepted unless other test review steps raise serious doubts about its validity.

10.2.8.2.4   If the PMSD measured for the test exceeds the upper PMSD bound variability criterion in Table 6, then one of the following two cases applies (10.2.8.2.4.1, 10.2.8.2.4.2).

10.2.8.2.4.1   If toxicity is found at the permitted receiving water concentration (RWC) based upon the value of the effect concentration estimate (NOEC or LOEC), then the test shall be accepted and the effect concentration estimate may be reported, unless other test review steps raise serious doubts about its validity.

10.2.8.2.4.2   If toxicity is not found at the permitted RWC based upon the value of the effect concentration estimate (NOEC or LOEC) and the PMSD measured for the test exceeds the upper PMSD bound, then the test shall not be accepted, and a new test must be conducted promptly on a newly collected sample.

10.2.8.2.5   To avoid penalizing laboratories that achieve unusually high precision, lower PMSD bounds shall also be applied when a hypothesis test result (e.g., NOEC or LOEC) is reported. Lower PMSD bounds, which are based on the 10th percentiles of national PMSD data, are presented in Table 6. The 10th percentile PMSD represents a practical limit to the sensitivity of the test method because few laboratories are able to achieve such precision on a

regular basis and most do not achieve it even occasionally. In determining hypothesis test results (e.g., NOEC or LOEC), a test concentration shall not be considered toxic (i.e., significantly different from the control) if the relative difference from the control is less than the lower PMSD bounds in Table 6. See USEPA, 2000b for specific examples of implementing lower PMSD bounds.

10.2.8.3   To assist in reviewing within-test variability, EPA recommends maintaining control charts of PMSDs calculated for successive effluent tests (USEPA, 2000b). A control chart of PMSD values characterizes the range of variability observed within a given laboratory, and allows comparison of individual test PMSDs with the laboratory's typical range of variability. Control charts of other variability and test performance measures, such as the MSD, standard deviation or CV of control responses, or average control response, also may be useful for reviewing tests and minimizing variability. The log of PMSD will provide an approximately normal variate useful for control charting.

TABLE 6.   VARIABILITY CRITERIA (UPPER AND LOWER PMSD BOUNDS) FOR SUBLETHAL HYPOTHESIS TESTING ENDPOINTS SUBMITTED UNDER NPDES PERMITS.[1]

Test Method	Endpoint	Lower PMSD Bound	Upper PMSD Bound
Method 1006.0, Inland Silverside Larval Survival and Growth Test	growth	11	28
Method 1007.0, *Mysidopsis bahia* Survival, Growth, and Fecundity Test	growth	11	37

[1]   Lower and upper PMSD bounds were determined from the 10th and 90th percentile, respectively, of PMSD data from EPA's WET Interlaboratory Variability Study (USEPA, 2001a; USEPA, 2000b).